

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 11, 2003, 20:06:42 ; Search time 2303 Seconds
(without alignments)
3866.895 Million cell updates/sec

Title: US-09-308-397-2

Perfect score: 1518

Sequence: 1 MTKTAFLFAGGQAYLGNWR.....QTAHLAHVEDQASIVALLEK 306

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09308397/runat.06062003.112348.28873/app.query.fasta_1.455
-DB=GenEmbl -QMT=fastap -SUFFIX=rce -MINMATCH=0.1 -LOOPEXT=0 -LCOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR.SCORE=500 -THR.WAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09308397@cgn.1.1.1652 @runat.06062003.112348.28873 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPEXT=0.5 -FGAPOPEXT=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

RESULT 1

29: em.vi:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pln:*
35: em.htg.rtd:*
36: em.htg.nam:*
37: em.htg.vrt:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB ID	Description
		Match	Length	%		
1	1518	100.0	921	6	AX194049	Sequence
2	1518	100.0	11523	1	AF197933	Streptoco
3	1518	100.0	12039	1	AE007354	Streptoco
C 4	1518	100.0	19702	6	BD003687	Polynucle
5	1518	100.0	111135	2	SPNEU1906	Streptoco
6	1515	99.8	945	6	AX194275	Sequence
7	1515	99.8	10925	1	AE008418	Streptoco
8	1048	69.0	10701	1	AE006311	Lactococc
C 9	1036	68.2	10909	1	AE006603	Streptoco
C 10	1036	68.2	50416	1	AE014165	Streptoco
C 11	1030	67.9	14456	1	AE010088	Streptoco
12	858	56.5	1196	6	AR193756	Sequence
C 13	775.5	51.1	11845	1	AE013105	Thermcana
C 14	710.5	46.8	13165	1	AE007854	Clostridi
15	703.5	46.3	29850	1	AP003389	Clostridi
16	669.5	44.1	27779	1	BSY13937	Bacillus su
17	669.5	44.1	208780	1	BSUB0009	Bacillus su
C 18	657.5	43.3	14530	1	AE004276	Vibrio ch
C 19	654	43.1	303249	1	AP001515	Bacillus
20	643.5	42.4	1440	1	ECFABHDG	E.coli fabh
21	643.5	42.4	1608	1	ECOFABD	Escherichia
22	643.5	42.4	13051	1	AE000210	Escherich
23	643.5	42.4	16379	1	D90745	Escherichia
24	640.5	42.2	12890	1	AE005319	Escherich
25	640.5	42.2	222605	1	AP002555	Escherich
26	636.5	41.9	2678	1	BSU59433	Bacillus su
C 27	636	41.9	10857	1	U32701	Haemophilus
28	628.5	41.4	13099	1	AE013778	Yersinia
29	628.5	41.4	193050	1	AJ414149	Yersinia
30	627.5	41.3	950	6	AX432728	Sequence
31	627.5	41.3	960	6	AX432671	Sequence
32	626	41.2	10684	1	AE006228	Pasteurel
33	616.5	40.6	4430	1	AF044668	Salmonell
34	616.5	40.6	20097	1	AE008752	Salmonell
35	612.5	40.3	254050	1	AL627269	Salmonell
36	611.5	40.3	6394	1	AB021978	Vibrio ma
37	598	39.4	193050	1	AL646062	Ralstonia
38	594.5	39.2	14103	1	AE000712	Aquifex a
39	589.5	38.8	2301	6	AX416528	Sequence
C 40	588.5	38.8	347050	1	AL591981	Listeria
C 41	586	38.6	313450	1	AL596170	Listeria
42	586	38.6	319630	6	AX413016	Sequence
C 43	586	38.6	349980	6	AX417046	Sequence
44	578	38.1	9922	1	AE012960	Chlorobiu
45	572.5	37.7	11361	1	AE011741	Xanthomon

ALIGNMENTS

AX194049
LOCUS
DEFINITION
Sequence 30 from Patent WO0149721.
ACCESSION
AX194049
VERSION
AX194049.1
KEYWORDS
GI:15211644
SOURCE
Streptococcus pneumoniae.
ORGANISM
Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 921)
AUTHORS
Dougherty,T.J., Pucci,M.J., Dougherty,B.A., Davison,D.B.,
Brucoleri,R.E. and Thanassi,J.A.
TITLE
Novel bacterial genes and proteins that are essential for cell
viability and their uses
JOURNAL
Patent: WO 0149721-A 30 12-JUL-2001;
Bristol-Myers Squibb Co. (US)
FEATURES
Location/Qualifiers
1..921
/organism="Streptococcus pneumoniae"
/db_xref="taxon:1313"
BASE COUNT 246 a 193 c 232 g 250 t
ORIGIN

Alignment Scores:
Pred. No.: 1,44e-109 Length: 921
Score: 1518.00 Matches: 306
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-308-397-2 (1-306) x AX194049 (1-921)

QY 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
Db 1 ATGACTAAACAGCCCTTTTATTTCGTGGTCAAGGTGCCAGGTATCTAGGATGGAGCGG 60
QY 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
Db 61 GATTTCATGATCAGTATCCGATTGTTAAGAACAAGATTGATCGAGCGAGTCAGGTGCTC 120
QY 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60
Db 121 GGTATTGATTTGCGTTATCTCATCGATACGAGAGAGACAAACTCAATCAGACCCGCTAT 180
QY 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGlyLysGly 80
Db 181 AGCAACACAGCCATCTTAGCACATTCGGTTGCTATCTACCCTTTATTCGAAGAAAGGCG 240
QY 81 TyrGlnProAspMetValAlaGlyLysSerLeuGlyGlyTyrSerAlaLeuValAlaSer 100
Db 241 TATCAGCCTGATATGTCGCTGGTTGTCCTTGAGAAATACCTCTGCTTGGTGCAAGC 300
QY 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
Db 301 GCGCCTTGGATTTTCAGATGCGGTTGCGCTTGGTAGCTTAAGCGTGGAGCCCTATTATGAA 360
QY 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
Db 361 GAAGCGGCTCCTGCTGACTCTGGCAGATGAGTAGCAGTTCTCAATACGCCAGTAGAGGTC 420
QY 141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
Db 421 ATTGAAGAAGCCGTGTCAAAAAGCTCTCTGAACCTTGGAGTGGTTTACTCCAGCCCACTATAC 480
QY 161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeu 180
Db 481 ACACCTGCACAATCGCTCATTCCTGGAGAGAGTGGTTGCAGTTGATCGAGCGGTTGACCTT 540
QY 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheIsthr 200
Db 541 TTGCAAGAAGCAGGTGCAACCGTTGATTCCTTTAAGGTGTCAAGGTGTCAAGTCCCTTTACACC 600

Qy	201	AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer	220
Db	601	GCTCTCCTTCAGGCAGCTAGCCGAAACTAGCTGAAACTCTAGCTCAGGTAAAGTTTTTCA	660
Qy	221	AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle	240
Db	661	GAITTTACTTGTCCCTAGTCGGCNAATACAGAGCTGCTGTGATGCAAAAGAGACATT	720
Qy	241	AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheThrGluSerIleGlyVal	260
Db	721	GCTCAGCTCTAGCGCTGAGTCAAGGAACCCGCTCGTTTCTATGAAAGTATTGGGGTC	780
Qy	261	MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly	280
Db	781	ATCCAGAAGCAGGCATCAAGCACTTATCGAGATTGACCGGGGAAGTCTTGTCAAGT	840
Qy	281	PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu	300
Db	841	TTTGTGTAATAAATAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTA	900
Qy	301	ValAlaLeuLeuGluLys 306	
Db	901	GTAGCACTTTTAGAAAAA 918	
RESULT 2			
AF197933			
LOCUS		11523 bp DNA linear BCT 10-AUG-2000	
DEFINITION		Streptococcus pneumoniae fab gene cluster, complete sequence.	
ACCESSION		AF197933	
VERSION		AF197933.1 GI:9789228	
KEYWORDS			
SOURCE		Streptococcus pneumoniae.	
ORGANISM		Streptococcus pneumoniae	
REFERENCE		Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
AUTHORS		Streptococcus.	
TITLE		1 (bases 1 to 11523)	
JOURNAL		Heath, R.J. and Rock, C.O.	
MEDLINE		A tricolous-resistant bacterial enzyme	
PUBMED		Nature 406 (6792), 145-146 (2000)	
REFERENCE		20365714	
AUTHORS		10910344	
TITLE		2 (bases 1 to 11523)	
JOURNAL		Heath, R.J. and Rock, C.O.	
SC		Submitted (25-OCT-1999) Department of Biochemistry, St Jude	
FEATURES		Children's Research Hospital, 332 North Lauderdale Street, Memphis,	
source		TN 38105, USA	
gene		Location/Qualifiers	
CDS		1. .11523	
		/organism="Streptococcus pneumoniae"	
		/strain="R6"	
		/db_xref="taxon:1313"	
		210. .1184	
		/gene="fabH"	
		210. .1184	
		/gene="fabH"	
		/note="FabH; KASIII; acetoacetyl-ACP synthase; by	
		similarity to FABH_ECOLI (39% identity)"	
		/codon_start=1	
		/transl_table=11	
		/product="beta-ketoacyl-ACP synthase III"	
		/protein_id="AAF9821.1"	
		/db_xref="GI:9789229"	
		/translations="MAFAKISQVHYPEQVTVNHDIAQMTNDENISSRTGIQRORH	
		ISRTGSTSLATVAKVAKAGITGEELFIILATITPDSMPSTAAHQVANTGANK	
		APAFDITACSGVFVALSTAEKFIASGRFQKGLVIGSETLSKAVDSMDRSTAVLFGDC	
		AGVFLLEASRQEFPLAESLNSDGRSECTLYGHSLHSPFSDQESADSLKMDGRTVFE	
		DPAIRDVAKSIKOTIDESPIEVDLDYLLHQANDRLDKMARKIGVDRAKLPANMME	
		YGNSTAASPIILLSECVQGLIPDGSQVLLSGFGGLTGWITDILII"	
gene		1244. .1468	
CDS		/gene="acpP"	
		1244. .1468	

/gene="acpp"
 /note="ACP; by similarity to ACP_ECOLI (38% identity)";
 /codon_start=1
 /transl_table=11
 /product="acyl carrier protein"
 /protein_id="AAF98272.1"
 /db_xref="GI:9789230"
 /translation="NAVEKQVQELIIVELGKDAEVTLESTFDDLDADSLDLFQVISE
 LEDAFDIOIEARNDUKTVGDLVAIVVEQAK"
 1587..2561
 /gene="fabk"
 1587..2561
 /gene="fabk"
 /note="FabK; NADH-dependent, FAD-containing enoyl-ACP
 reductase"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="trans-2-enoyl-ACP reductase II"
 /protein_id="AAF98273.1"
 /db_xref="GI:9789231"
 /translation="MKRTITELTKIDYPIFOGMAWADGLAGAVSKAGGLGIIGG
 NAPKSVKANIDKISLTDKPGVNMILLSPVEDIVLVEEGVKVVTTCAGNPSKY
 MERFHEAGLIVPVPSVALAKMEKIGADAVIAGMEAGGHIGKLTITMLVROVAIA
 ISIPVIAAGGIDGGAAGFMAEAVOVGTRFVAKESNAHPNYKEKILKARDIDT
 TISAQFHGAHRAIKNQTRDFELAKDAPKQDDPDLEIFQMGAGALAKAVHGDVD
 GGSVMAGQIAGLVEETAEELIKLYYGAAKKIOEEASRWTVGVVND"
 2554..3474
 /gene="fabD"
 2554..3474
 /gene="fabD"
 /note="FabD; by similarity to FABD_ECOLI (42% identity);
 MCAT"
 /codon_start=1
 /transl_table=11
 /product="malonyl-CoA:ACP transacylase"
 /protein_id="AAF98274.1"
 /db_xref="GI:9789232"
 /translation="MTKTAFLFAGGQAQVYLGMRDFYQYPIVKTETIDRASQVLGYDL
 RYLITTEEDKLNQTRYPAILATSAIYRLGQYQDMVAGLSGEYSALVASGA
 LDFEAVALLKAPMEAEAPADSKRMVAVLNTPEVIEACQKASGLGVTFPANTY
 TPAQVIVAGEVAVDRVAVELLOEAGAKRLIPLKVGSPHTALLPASOKLAETLAQVS
 RSDFTCLVGNTEAAVMOKEDIAQLLTRQVKEPVRFYESIGVMQEAQISNFIETGPQK
 VLSGFVKIDQTAHLAHVEDQASLVALLEK"
 3508..4239
 /gene="fabG"
 3508..4239
 /gene="fabG"
 /note="FabG; by similarity to FABG_ECOLI (44% identity)";
 /codon_start=1
 /transl_table=11
 /product="beta-ketoacyl-ACP reductase"
 /protein_id="AAF98275.1"
 /db_xref="GI:9789233"
 /translation="MKLEHKNFITGSSRGIGLTAHKFAQAGANIVILNSRGAISEEL
 LAEFNNGYKVPISGDVDPADAKRMDQIAELGSDVLVNNAGITQDTLMUKMTE
 ADFEKLVTNLGAFNMQSVLKPWKAREGALINWSSVGLMGNGIQAAYAAKAGL
 IGFTKSVAREVASRIRNVNVIAPCMIESDHTALLSDKIKETLAQIPKKEFGQAQEA
 DUTFLAGQDLYTGQVVAIDGGLSM"
 4261..5496
 /gene="fabF"
 4261..5496
 /gene="fabF"
 /note="FabF; KASII; oxoacyl-ACP synthase II; by similarity
 to FABF_ECOLI (43% identity) and FABF_ECOLI (33%
 identity)";
 /codon_start=1
 /transl_table=11
 /product="beta-ketoacyl-ACP synthase II"
 /protein_id="AAF98276.1"
 /db_xref="GI:9789234"
 /translation="MKLNRVVTVGTGVTSPIGNTPEEFWNSLATKIGIGITKFDHS
 DFDVHNAEIQDFPDKPYEVKDTNRFDNISLYALAAQEAHVNHANLDVEALNRDRFG

VIVASGIGGKEIQDLVLRHEKPKRVKPMTLPRALPNMAGNVAMRFGANGVCXKSI
 NTACSSNDAGAPRSIKFQDVMVAGTASITPPALAGQALTAJLSTTDDPTRA
 SIFPDKDRNGFVNGSGMLVLESLEHAKRGATILAEVYGVNCTGDVHMTSPHEG
 QGAIKAKLAEAEISPEQVAVYVNAHGTSTPANEGESGAIVVLGKVFVSSTKSF
 TGHLLAGAGAVEAIVTEAMRNFVEMTGTSEVSDIIEANVVYVIGGLEKEIPYALSN
 TFGFGGHNALAFKWRNR"
 5499..5984
 /gene="accB"
 5499..5984
 /gene="accB"
 /note="BCCP; by similarity to BCCP_ECOLI (37% identity);
 subunit of acetyl-CoA carboxylase"
 /codon_start=1
 /transl_table=11
 /product="biotin carboxyl carrier protein"
 /protein_id="AAF98277.1"
 /db_xref="GI:9789235"
 /translation="MNLNDIKLMTQFDOSILREERSYKNGTDELQSKNEARVPVEVA
 TVQAPAPVILATSPVAPTVAEVPAPAEASVATEGNLVSPLGVVYLAAGP
 DKPAFTVGDVSKGQGLTVIIEAMKVMNEIPAKDGVTEILLVSNEEMVEFGKGLVRI
 K"
 5981..6403
 /gene="fabZ"
 5981..6403
 /gene="fabZ"
 /note="FabZ; by similarity to FABZ_ECOLI (46% identity)";
 /codon_start=1
 /transl_table=11
 /product="beta-hydroxyacyl-ACP dehydratase"
 /protein_id="AAF98278.1"
 /db_xref="GI:9789236"
 /translation="MIDIGIKELPHRPMLLVDRVLSVEDTIVAIGNVTINEPFF
 NGHFPQYVNVGVVINEALQATAGVLELSPENKGLVFIAGMDKVKFKQVVGQDL
 VMTATFVRRGTIAVVEAKAEVNDKLAASGTLTFAIGN"
 6415..7782
 /gene="accC"
 6415..7782
 /gene="accC"
 /note="AccC; by similarity to ACCC_ECOLI (53% identity)";
 /codon_start=1
 /transl_table=11
 /product="acetyl-CoA carboxylase biotin carboxylase
 subunit"
 /protein_id="AAF98279.1"
 /db_xref="GI:9789237"
 /translation="MPKLLIANRGEIARIIPARELGIATVAVYSTADKEALHTLL
 ADEAVCTIGPKATESYLNINAVLSAAVLTEASAIHPFGFLSENSFPATCEVGVKX
 IGPSHVMWMDGKINARQMIKAGVFIIPSGDGEVHNSSEALIVAEKIGYVNLKAS
 AGGGKIGIRKVEKPDLLVSAPETASSEAKANYNGANGAYIERVIYPARHIEVILGDEH
 GHVILHGERDCSLQRNNQVLEBSPSIAIGKTLRHEIGAAVRAAFAVGVENAGTIEF
 LLDEASSNFYEMEMNTRVQVEHPVTFVSGVDIVKEQICIAAGQPLSVQEDIVLRGH
 AIECRINAENPAENFAPSQKLTNLYLPSGGVGLRVDSAVYPGYTPPYVDSMIKII
 VHGENFDALMKQALYELIEGTQTNADQLDLISDRNVIAGDIDTSELMETFLPK
 YQKE"
 7819..8685
 /gene="accD"
 7819..8685
 /gene="accD"
 /note="AccD; by similarity to ACCD_ECOLI (43% identity)";
 /codon_start=1
 /transl_table=11
 /product="acetyl-CoA carboxylase beta subunit"
 /protein_id="AAF98280.1"
 /db_xref="GI:9789238"
 /translation="MALFSKDKVIRINRNRVREKPAQPEVDELFSOCPGCKHTI
 YXDLGSEIRCPHCSYTERISAQERLALITDMCTFKELTGIESKDPHPGQYKKLIA
 SMREKGLHRAVVTGTALINGQVAGLIMDSNFMASMGTVGKILTRIFEYATVEKL
 PVVLTASGARMQCEGIMSLMQNAKISAAVHRNSNAGLFLTILTDPTTGGVGTASAM
 EGDIIILBPQSLVGPAGRRVIENTVRESUPEDFQKAEFFLLEHGFDAIVKRRDLPTDI
 ASLVRLHGSGSR"
 8682..9449
 /gene="acca"
 8682..9449

```
/gene="acca"  
/notes="ACCA, by similarity to ACCA_ECOLI (46% identity)"  
/codon_start=1  
/transl_table=11  
/product="acetyl-CoA carboxylase alpha subunit"  
/protein_id="AAF98281.1"  
/db_xref="GI:9789239"  
/translation="MNIAKIVREAREQSELTLDFATGTFDFEFLHGDHSFRDGGAV
```

Alignment Scores:

```
Pred. No.: 3,62e-108 Length: 11523  
Score: 1518.00 Matches: 306  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 1 Gaps: 0  
  
US-09-308-397-2 (1-306) x AF197933 (1-11523)  
  
QY 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20  
Db 2554 ATGACTAAACAGCCCTTTATTTGCTGCTCAAGTGCCAGTATCTAGGGATGGACGG 2613  
QY 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40  
Db 2614 GATTTCTATGATCAGTATCGGATTTGTTAAAGAAACGATTGATCGAGCGACCGGTGCTC 2673  
QY 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60  
Db 2674 GGTATGATTTGGTTATCTCATCGATACGGAAGAACAACTCAATCAGACCCGCTAT 2733  
QY 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnLysGly 80  
Db 2734 ACCCAACCGACCTTCTAGCAGTTCGGTTGCTATCTACCGTTTATTGCAAGAAAGGCG 2793  
QY 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyValTyrSerAlaLeuValAlaSer 100  
Db 2794 TATCAGCCTGATATGTTGCTGCTGTTGCTCTTGGAGATATCTGCTTGGTGGCAAGT 2853  
QY 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120  
Db 2854 GCGCGCTTGATTTTGAAGATGGGTGCTGCTGTTGATGCTTAAGCGTGGAGCCCTATATGAA 2913  
QY 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140  
Db 2914 GAAGCGCTCTCTGCTGCTGCTGCAAGATGTTAGCAGTTCTCAATACGCCAGTACAGTTC 2973  
QY 141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaSerThrAsn 160  
Db 2974 ATTGAAGAGCCCTGTCAAAAGCTTCTGAACCTGGAGTGGTGTACTCCAGCCCACTATAAC 3033  
QY 161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeu 180  
Db 3034 ACACCTGCACAAATCGTCATTTGCTGAGAGAGTGGTTGCGAGTGTATGACGCGGTGAACCT 3093  
QY 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheIleThr 200  
Db 3094 TTGCAAGAGCAGGTGCCAAACGCTTGATTCCTCTTAAGGTGTCAAGTCCCCTTCACACC 3153  
QY 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrIleAlaGlnValSerPheSer 220  
Db 3154 GCTCTCCTTGAACCTGCTAGCCAGAACTAGCTGAAACTCTAGCTCAGTAAGTTTTCATCA 3213  
QY 221 AspPheThrCysProLeuValGlyValAsnThrGluAlaAlaValMetGlnLysGluAspIle 240  
Db 3214 GATTTTACTTGTCCCTAGTCCGCAATACAGAAGTGTCTGTATGCAAAAGAGGACATT 3273  
QY 241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260  
Db 3274 GCTCAGCTCTGACGCGTCAAGTCAAGGAAACCCGCTTCTTCTATGAAAGTATGCGGTC 3333  
QY 261 MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280
```

```
Db 3334 ATGCAAGACAGGCATACCAACTTTATCGAGATTGAGCGGGAAGTCTTTGTTCAGGT 3393  
QY 281 PheValLysLysLeuAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300  
Db 3394 TTTGTTAAAAAATTGATCAAACTGCTCTTACCTTACCTCATGTGGAAGATCAAGCGAGTTTA 3453  
QY 301 ValAlaLeuLeuGluLys 306  
Db 3454 GTAGCACTTTTAGAAAA 3471  
  
RESULT 3  
AE007354 12039 bp DNA linear BCT 31-AUG-2001  
LOCUS Streptococcus pneumoniae TIGR4 section 37 of 194 of the complete  
DEFINITION genome.  
ACCESSION AE007354 AE005672  
VERSION AE007354.1 GI:114971888  
KEYWORDS Streptococcus pneumoniae TIGR4.  
SOURCE Streptococcus pneumoniae TIGR4  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 12039)  
AUTHORS Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D.,  
Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J.,  
Durkin, A.S., Winn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D.,  
Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D.,  
Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L.,  
McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T.,  
Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O.,  
Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and  
Fraser, C.M.  
TITLE Complete genome sequence of a virulent isolate of Streptococcus  
pneumoniae  
JOURNAL Science 293 (5529), 498-506 (2001)  
MEDLINE 21357209  
PUBMED 11463916  
REFERENCE 2 (bases 1 to 12039)  
AUTHORS Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D.,  
Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J.,  
Durkin, A.S., Winn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D.,  
Umayam, L.A., White, O., Lewis, M.R., Radune, D., Holtzapple, E.,  
Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A.,  
Feldblyum, T.V., Angiuoli, S., Gesuwan, P., Hickey, E.K., Holt, I.E.,  
Loftus, B.J., Ujwal, M.L., Yang, F., Smith, H.O., Venter, J.C.,  
Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.  
Direct Submission  
TITLE Submitted (29-JUN-2001) The Institute for Genomic Research, 9712  
JOURNAL Center Dr, Rockville, MD 20850, USA  
MEDLINE  
PUBMED  
FEATURES  
source 1. 12039  
/organism="Streptococcus pneumoniae TIGR4"  
/strain="TIGR4"  
/db_xref="taxon:170187"  
90. .1064  
/gene="SP0419"  
90. .1064  
/gene="SP0419"  
/codon_start=1  
/transl_table=11  
/product="enoyl-(acyl-carrier-protein) reductase"  
/protein_id="AAK74582.1"  
/db_xref="GI:114971889"  
1057. .1977  
/gene="SP0420"  
1057. .1977  
/gene="SP0420"  
  
gene  
CDS  
  
gene  
CDS
```

```

/Note="identified by match to TIGR protein family HMM
TIGR00128"
/codon_start=1
/transl_table=11
/product="malonyl CoA-acyl carrier protein transacylase"
/protein_id="AAK74583.1"
/db_xref="GI:14971890"
/translation="MKTAFLEFAGQAGYLCGRDYPDQPIVKETIDRASQVLGYDL
RYLIDTEEDLNQTRYTPQPAITLVAIYELLQEKGYQPMVAGLSGBYSALVAGSA
LDFEDAVLAKGAYMEAPADSGKQWAVLNPVEVIEACQKASELGVVTPANYN
FSPQITAGVAVDRAVELLQEGAKRLIFLKVSGPFHTALLEPASOKLAETLAQVS
TSAQTCVLGNTGTEAAVMQKEDIQALLTROKVEPVRVESIGNVQKEAGISNFIHGPK
VLSGEVKIKDQTAHLAHEVDQASLVALLEK"
2011..2742
/gene="SP0421"
/Note="SP0421"
/Note="SP0421"
/codon_start=1
/transl_table=11
/product="3-oxoacyl-(acyl-carrier protein) reductase"
/protein_id="AAK74584.1"
/db_xref="GI:14971891"
/translation="MKLEHKNIFITGSSRIGIGIAIAHKFAQAGANIVLSRGAISBEL
LAEPSTGKIVPISGDSVDFADAKRMIDQAIAGLSGVDVLYNNAGITQDTLMKMT
ADEPKVLKVLGTAFNTQSVLPKMKAREGAIINMSVVLGNIGQANYAASKAGL
IGFTKSVAREVARNIRNVIAFGMIESDMTAILSDKIKEATLAQIPMKFEFGQAEQVA
DLTVLAGQDYLTSQVVAIDGGLSM"
2764..3999
/gene="SP0422"
/Note="SP0422"
/Note="SP0422"
/codon_start=1
/transl_table=11
/product="3-oxoacyl-(acyl-carrier-protein) synthase II"
/protein_id="AAK74585.1"
/db_xref="GI:14971892"
/translation="MKLNRVVTVGYVTSPIGNTBPFNWSLATGKIGIGITKFDHS
DFDVAAEIQDFPFKYFVKQDNTDFNYSLYAAQAVNHANLDVLEALNRDRFG
VIVASGIGKEIDBDVLHKEGPKVKMTPLKALPNVNSGAMVRFGANGVCKSI
NTACSSNDIGAIDAFRIKFGQDVLGVGTGEASITPFAAGFALTALSTLTDPTRA
SIPDKORNGPVMGEGGMLVLESHAEKRGATILAEVVGNGTCDVHMTSPHPEG
OGAIKATLALAEAEISPEQAVYNAHGTSTPANEGESCAIVAVLKGKVPVSTKSP
TGHLLGAGAVEALVTIEMAHNFVPMTAGTSEVSDYIEANVVYGGGLEKEIPIYALSN
TFPGGHNVAIAFRWENR"
4002..4487
/gene="SP0423"
/Note="SP0423"
/Note="SP0423"
/codon_start=1
/transl_table=11
/product="acetyl-CoA carboxylase, biotin carboxyl carrier
protein"
/protein_id="AAK74586.1"
/db_xref="GI:14971893"
/translation="MNLDIKDMLTQDQSSLRBFYKNGTDELQFSKNEARVPVPA
TQVAPVLPATPSVAPTSAPAEVAPAEASVATEGNLVSPLGVVYLAAGP
DKPAFVTVGDSVKKGQTLVITIAMKVMNEIPAPKDGVTVELVSNNEEMVEFGKGLVRI
K"
4484..4906
/gene="SP0424"
/Note="SP0424"
/Note="SP0424"
/codon_start=1
/transl_table=11
/product="(3R)-hydroxymyristoyl-(acyl-carrier-protein)
dehydratase"
/protein_id="AAK74587.1"
/db_xref="GI:14971894"
/translation="MIDIQIKKEALPHRYPMILLVDRVLEVEDTIVAIKNVTINEPFF
NGHPFPVPMGVVMEALQAGVLESPENKGLVFIYAGMDKVKRKKQVVPVGGQL
VMTATFYKRGTTIAVWEAKAEVDGKLAASGTLTFAIGN"

```

```

gene
CDS
4918..6285
/gene="SP0425"
/Note="SP0425"
/Note="SP0425"
/Note="identified by match to PFAM protein family HMM
PF00289"
/codon_start=1
/transl_table=11
/product="acetyl-CoA carboxylase, biotin carboxylase"
/protein_id="AAK74588.1"
/db_xref="GI:14971895"
/translation="MFRKILIANRGEIAVRIIRAARELGIAIVAVYSTADKEALHTLL
ADEAVICGPGKATSYLNINAVLSAAVLTEAEAIHPGFLSENSKPMATMCEEVGKIF
IGPSGVMDMGKDKINAPAOIKAGVPVPCSDGEVHNSBEALIVAEKIGYPMVKAS
AGGKGKIRKVEKPDILVSAPEATSSAKANYGNGAMVIERVYPAHIEVQLLGDH
CHVHLGERSCLORNNOKVLEESPSAIGKTLRHEIGAAAVRAAEFVGYENAGTIEF
LLDEASNFYFMEQNRVQVBEHPVTEFVSGVDIVKEQICIAAGQPLSVKQSDIVLRGH
AIECRINAENAFNFPAPGPKITNLYLPSGGVGLRVDSAVPYGTIPPYDSMIAKII
VHGNRFDALMKMORALYELIEGVQTNADQLDLISDRNVLAGDYDTSFLMETFLPK
YQKEK"
6322..7188
/gene="SP0426"
/Note="SP0426"
/Note="SP0426"
/Note="identified by match to TIGR protein family HMM
TIGR00515"
/codon_start=1
/transl_table=11
/product="acetyl-CoA carboxylase, carboxyl transferase,
beta subunit"
/protein_id="AAK74589.1"
/db_xref="GI:14971896"
/translation="MAFSKKDKYIRINPNSVREKPOAKPEVDELFSCPGCKHTI
YKDLGSRICPHGCSYTPRIQAORLALTIMDGTPELFTGIESKDLPHFGYQKLLA
SMREKTHGEAVVTGTLIKGTVALGMDNFMASMGTVVGEKITLFEVATVEKL
PVVLTASGARMQEGIMSLMQAKISAAVKEHNSAGLYLTITLDTPTGGVATSFAM
EGDILAPQPSLVFAGRRVIENTVRESLPDFQKAEFLLEHGFVDALVKRDLPDTI
ASLVLHGGSPR"
7185..7952
/gene="SP0427"
/Note="SP0427"
/Note="SP0427"
/Note="identified by match to TIGR protein family HMM
TIGR00513"
/codon_start=1
/transl_table=11
/product="acetyl-CoA carboxylase, carboxyl transferase,
alpha subunit"
/protein_id="AAK74590.1"
/db_xref="GI:14971897"
/translation="MNIKIVREAREQSRLTLDTFGTDFEFTQLHGRSPRDDGAV
VGGIWLGDQAVTVVGIQKSLQNLKRNFGQHPGEGYKALRMKQAEKFGPVT
FINTAGYGVGAEEGGEATARNMEMSCLKVPIITAIIGEGSGGALALAVADRV
WLENSIVAILSPGFSASILMKDGTTRAMEALMKITSHELLEMDVVDKVISEIGLSS
KELIKSVKKEIQTELARLSQKPLEBEEYQRFKY"
7963..8070
/gene="SP0428"
/Note="SP0428"
/Note="SP0428"
/Note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK74591.1"
/db_xref="GI:14971898"
/translation="MRLRSHFLMKWVIOEKLPEVEMGYKIDKRNIL"

```

Alignment Scores:

Pred. No.:	3,836-108	Length:	12039
Score:	1518.00	Matches:	306
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0

```

DB: 1 Gaps: 0
US-09-308-397-2 (1-306) x AB007354 (1-12039)
QY 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlnValaGlnTyrLeuGlyMetGlyArg 20
Db 1057 ATGACTAAACAGCCCTTTTATTGCTGCTCAAGTGGCCAGTACTAGGATGGACGG 1116
QY 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
Db 1117 GATTTCATGATCAGTATCCGATTTGTTAAAGAACGATTCGAGCCGAGCCAGGTGCTC 1176
QY 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60
Db 1177 GGTATGATTTGGTATCTCATCGATCGAAGAGACAACTCAATCAGACCCGCTAT 1236
QY 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLysGly 80
Db 1237 AGCGAACACGCACTTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAGGCG 1296
QY 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
Db 1297 TATCAGCCGTATGATGTTGCTGCTGTTGCTCTTGAGAAATATCTGCTTGGTGCAAGT 1356
QY 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
Db 1357 GCGCCCTTGATTTGAAGATGCGTTCGCTTGGTAGCTTAAGCGTGGAGCCCTATATGGA 1416
QY 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
Db 1417 GAAGCGGCTCTGCTGACTCTCGCAAGATGGTAGAGTTCTCAATACGCCAGTAGAGTGC 1476
QY 141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
Db 1477 ATTGAAGAGCCCTGCAAAAGCTTCTGAACCTGGAGTGGTTACTCCAGCCAATATAAC 1536
QY 161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAlaAspArgAlaValGluLeu 180
Db 1537 ACACCTTGCAAAATGCTCATCTCGAGAAAGTGGTTCAGATTGATCGAGCCGTTGAACCT 1596
QY 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr 200
Db 1597 TTGCAAGAGCAGGTGCCAAGCTTGATTCCTCTTAGGTGTCAAGTCCCTTTACACACC 1656
QY 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
Db 1657 GCTCTCCTTGAACTGCTAGCCAGAACTAGCTGAACACTCTAGCTCAGGTAAAGTTTTC 1716
QY 221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle 240
Db 1717 GATTTTACTTGTCCCTAGTCGCAATACAGAACTGCTGTGTGATGCCAAGAGAGACAT 1776
QY 241 AlaGlnLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
Db 1777 GCTCAGCTCTTACCGGCTCAGGTCAAGAAACCCGTTGCTTCTATGAAGATATTGGGTC 1836
QY 261 MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280
Db 1837 ATCAAGAGAGCAGCATAGCAACTTTATCGAGATTGGACCCGGGAAAGTCTTGTGAGT 1896
QY 281 PheValLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
Db 1897 TTGTGTTAAAAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTA 1956
QY 301 ValAlaLeuLeuGluLys 306
Db 1957 GTAGCACTTTTAAAAA 1974
RESULT 4
BD003687/c 19702 bp DNA linear PAT 31-JAN-2002
LOCUS BD003687
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003687

```

```

VERSION BD003687.1 GI:18631648
KEYWORDS JP 2001501833-A/7.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 19702)
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Bara, S.C.,
Pannon, V., and Dougherty, B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 7 13-FEB-2001;
HUMAN GENOME SCIENCES INC
COMMENT OS Unidentified
PN JP 2001501833-A/7
PD 13-FEB-2001
PF 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI
STEVEN C BARASH,
PI MICHAEL FANNON, BRIAN A DOUGHERTY
PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19,
PC C12N1/21,
PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC
G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..19702
FT /organism='Unidentified'.
FEATURES
source Location/Qualifiers
1..19702
/organism='unidentified'
/db_xref='taxon:32644'
BASE COUNT 5736 a 4548 c 3728 g 5690 t
ORIGIN
Alignment Scores: 7.17e-108 Length: 19702
Pred. No.: 1518.00 Matches: 306
Score: 1518.00
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-308-397-2 (1-306) x BD003687 (1-19702)
QY 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlnValaGlnTyrLeuGlyMetGlyArg 20
Db 17149 ATGACTAAACAGCCCTTTTATTGCTGCTCAAGTGGCCAGTACTAGGATGGACGG 17090
QY 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
Db 17089 GATTTCTATGATCAGTATCCGATTTGTTAAAGAACGATTCGAGCCGAGCCAGGTGCTC 17030
QY 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60
Db 17029 GGTATGATTTGGTATCTCATCGATACGGAAGAACAACTCAATCAGACCCGCTAT 16970
QY 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLysGly 80
Db 16969 ACCCAACCCAGCCATTCAGCACTTCGGTTGCTATCTACCGTTATTGCAAGAAAGGCG 16910
QY 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
Db 16909 TATCAGCCGTATGATGTTGCTGCTGTTGCTCTCTTGAGAAATATCTCTGCTTGGTCAAGT 16850
QY 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
Db 16849 GCGCCCTTGATTTGAAGATGCGGTTGCTTGGTAGCTGAAGCGTGAGCCCTATATGGA 16790
QY 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
Db 16789 GAAGCGGCTCTGCTGACTCTGCAAGATGGTAGCAAGTTCTCAATACGCCAGTAGAGTTC 16730

```

QY	141	IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn	160
Db	16729	ATTGAAGAAGCCTGTCACAAAAGCTTCTGAACCTTGGAGTGGTTACTCCAGCCAACTATTAAC	16570
QY	161	ThrProAlaGlnIleValIleAlaGlyGluValAlaValAlaValAspArgAlaValGluLeu	180
Db	16669	ACACCTGCACAAATCGTCATTCTCTGGAGAGTGGTTGCAGTTGATCGAGCGGTTGAACCT	16610
QY	181	LeuGlnGluAlaGlyAlaLysArgIleuLeuProLeuLysValSerGlyProPheHisThr	200
Db	16609	TTTGCAAGAAGCAGGTGCCAAAGCGCTGATTCCTCTTAAGGTGTCAAGTCCCTTTCCACACC	16550
QY	201	AlaLeuLeuGluProAlaSerGlnLysIleuAlaGluThrLeuAlaGlnValSerPheSer	220
Db	16549	GCTCTCCCTTGAACCTCTAGCCAGAAACCTAGCTGAACCTCTAGCTCAGTGAAGTTTTCAT	16490
QY	221	AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle	240
Db	16489	GATTTTACITGTCTCCCTAGTCGCAATACAGAAGCTGCTGTGATCAAAAAGAGGACATT	16430
QY	241	AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal	260
Db	16429	GCTCAGCTCTTGACGGCTCAGTCAAGCAACCGCTCGCTTCTATGAAGATATTGGGGTC	16370
QY	261	MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly	280
Db	16369	ATGCAGAGACGCGCATAGCAACTTATTCGAGATTGGACCGGGAAGATCTTGTCAGGT	16310
QY	281	PheValLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu	300
Db	16309	TTTGTGTAATAAATTCATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTA	16250
QY	301	ValAlaLeuLeuGluLys 306	
Db	16249	GTAGCACTTTTGAAGAAA 16232	
RESULT 5			
SPNEU1906			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			

```

Db      27318 TTTGTTAAAAAATTGATCAAACTGCTCACTTAGCTCATGTGAGATCAAGCGAGTTTA 27377
Qy      301 ValAlaLeuGluLys 306
Db      27378 GTAGCACCTTTTAGAAAAA 27395

RESULT 6
AX194275 AX194275 945 bp DNA linear PAT 15-AUG-2001
LOCUS Sequence 256 from Patent WO0149721.
ACCESSION AX194275
VERSION AX194275.1 GI:152111757
KEYWORDS Streptococcus pneumoniae.
SOURCE Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
ORGANISM Streptococcus.
REFERENCE 1 (bases 1 to 945)
AUTHORS Dougherty, T.J., Pucci, M.J., Dougherty, B.A., Davison, D.B.,
          Bruccoleri, R.E. and Thanassi, J.A.
TITLE Novel bacterial genes and proteins that are essential for cell
viability and their uses
JOURNAL Patent: WO 0149721-A 256 12-JUL-2001;
          Bristol-Myers Squibb Co. (US)
FEATURES
    source
        location/Qualifiers
            1..945
                /organism="Streptococcus pneumoniae"
                /db_xref="taxon:1313"
BASE COUNT 254 a 205 c 232 g 254 t
ORIGIN

Alignment Scores:
Pred. No.: 2,55e-109 Length: 945
Score: 1515.00 Matches: 305
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.67% Mismatches: 0
Query Match: 99.80% Indels: 0
DB: 5 Gaps: 0

US-09-308-397-2 (1-306) x AX194275 (1-945)

Qy      1 MetThrythrAlaPheLeuPheAlaGlyGlnGlyValaGlnTyrLeuGlyMetGlyArg 20
Db      1 ATGACTAAACACCCCTTTTATTGCTGTCAGGTCGCCAGTATCTAGGATGGACGG 60

Qy      21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
Db      61 GATTTCATCATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCGAGTCAGGTGCTA 120

Qy      41 GlyTyrAspLeuArgTyrIleuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60
Db      121 GGTATGATTGGTTGTTATCTCATCGATACGGAAGAACAACTCAATCAGACCCGCTAT 180

Qy      61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGlnGluLysGly 80
Db      181 ACGCAACCCGCACTTACGACCTTCGGTTGCTATCTACCGTTTATTGCAAGAAAAGGCG 240

Qy      81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
Db      241 TATCAGCCCTGATATGTTGCTGTTGTTCTCTTGGAGATATCTGCTTGTGGCAAGC 300

Qy      101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
Db      301 GCGCCCTGGATTGTTGAGATGCGGTTGCTTGTGAGTAAAGCTGAGCCTATATGNA 360

Qy      121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
Db      361 GAAGCGGCTCCTGCTGACTCTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGTGC 420

Qy      141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
Db      421 ATTGAAGACCCCTGTCAAAAGCTTCTGAACTTGGAGTGGTACTCCAGCCCACTATAC 480

```

```

Qy      161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeu 180
Db      481 ACACCTGCACAAATCGTCATTGCTGGAGAAGTGGTTGCAGTTGATCGAGCGTTGAACCTT 540

Qy      181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr 200
Db      541 TTSCAAGAAGCAGGTGCCAAACGCTTGATTCCTCTTAAGGTGTCCAGTCCCTTCACACC 600

Qy      201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
Db      601 TCTCTCTTGAACCTGCTAGCCAGAACTAGCTGAAACTCTGGCTCAGGTAAAGTTTTC 660

Qy      221 AspPheThrCysProLeuValGlyAsnThrGluAlaValMetGlnLysGluAspIle 240
Db      661 GATTTCATTGTCCTCTAGTCGCATACAGAAGCTGCTGTGATGCAAAAAGAGGACATT 720

Qy      241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
Db      721 GCTCAGCTCTTGACGCGTCAGTCAAGGAACCGCTTCGTTCTATGAAAGTATTGGGGTC 780

Qy      261 MetGlnGluAlaGlyIleSerAsnPheIleGlyProGlyLysValLeuSerGly 280
Db      781 ATGCAAGAAGCAGCATACAACTTTATCGAATTGGACCGGGGAAAGTCTTGTCAGGT 840

Qy      281 PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
Db      841 TTTGTTAAAAAATTGATCAAACTGCTCACTAGCTCATGTGAAGATCAAGCGAGTTTA 900

Qy      301 ValAlaLeuGluLys 306
Db      901 GTAGCACCTTTTAGAAAAA 918

RESULT 7
AE008418 10925 bp DNA linear BCT 13-SEP-2001
LOCUS Streptococcus pneumoniae R6 section 34 of 184 of the complete
DEFINITION genome.
ACCESSION AE008418 AE007317
VERSION AE008418.1 GI:15457935
KEYWORDS Streptococcus pneumoniae R6.
SOURCE Streptococcus pneumoniae R6
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1 (bases 1 to 10925)
AUTHORS Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,
          DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C.,
          Gilmour, R., Glass, J.S., Khoja, H., Kraft, A., Lagace, R.,
          LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,
          McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicas, T.I.,
          Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rockey, P.,
          Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,
          Zook, C., Baltz, R.H., Jaskunas, S.R., Rostek, P.R. Jr., Skatrud, P.L.
          and Glass, J.I.
TITLE Genome of the bacterium Streptococcus pneumoniae strain R6
JOURNAL J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21429245
PUBMED 11544234
REFERENCE 2 (bases 1 to 10925)
AUTHORS Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,
          DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C.,
          Gilmour, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R.,
          LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,
          McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicas, T.I.,
          Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rockey, P.,
          Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,
          Zook, C., Baltz, R.H., Jaskunas, S.R., Rostek, P.R. Jr., Skatrud, P.L.
          and Glass, J.I.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and
          Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
FEATURES Location/Qualifiers

```



```

source
1. .10925
/organism="Streptococcus pneumoniae R6"
/strain="R6"
/db_xref="taxon:171101"
complement(122..233)
/repeat_region
/rpt_family="BOX element"
/rpt_type="dispersed"
/rpt_unit="Box_CBA
264..1049
/gene="phaB"
/notes="spr0375"
264..1049
/gene="phaB"
/EC number="4.2.1.17"
/notes="Putative"
/codon_start=1
/transl_table=11
/label=spr0375
/product="Enoyl-CoA hydratase II"
/protein_id="AAK99179.1"
/db_xref="GI:15457936"
/transl_table=11
/translation="MEHIIYQLEDLAILNRPVANGPHIPMCREILLEALTAREEN
PAVHFLINANGKVSFGDLVEMKRAVEDDIPSLTKIAELVNTISYKIKQIAKPV
NEVDCVAGAAANMAVADPCLDKAKFIQAFVGLAPDAGIHLRSIGVTRAA
QAWTGALTAELALEGLVIRVSEAEKIKETREQLLKLRRASSNSYAAIKLVWES
QKDWQGYATLELNLOKSLAQTEDFEGVRAHSRRPKFIQR"
1143..1577
/gene="spr0376"
1143..1577
/gene="spr0376"
/notes="Probable transcriptional regulator"
/codon_start=1
/transl_table=11
/label=spr0376
/product="Conserved hypothetical protein"
/protein_id="AAK99180.1"
/db_xref="GI:15457937"
/transl_table=11
/translation="MDYQRIEVLTSIFNNVLVIBEVNLRGSRFKDISIKEMTIDVI
KAPDVPQSQKELMTVLTGTTSLNNLRKGYIERVSEQDRVVHLHLTKKGRLI
HRLKRFHKAVEXKIIDGMSEEEIIVMGKGLTNLYQFLEDLK"
1577..2551
/gene="fabH"
/notes="spr0377"
1577..2551
/gene="fabH"
/EC number="2.3.1.41"
/codon_start=1
/transl_table=11
/label=spr0377
/product="3-oxoacyl-acyl-carrier-protein synthase III"
/protein_id="AAK99181.1"
/db_xref="GI:15457938"
/transl_table=11
/translation="MAFAKISQVAHYVPEQVWNHDLAQIMDTNDEWISRTGIRQH
ISRTESDTEATEVAKLMKAGITGEELDFIILATITPDSMPSTAARVANGANK
AFAPDLAACSQGFPALESTAEKFIASGRFQGLVIGSETLSKAVDSDRSTAVLFDG
AGVLLBASQHEHLESLNSDGRSCLTYHSGHLSPFSDOESADSFLKMDGRTVF
DFAIRDVAKSIKOTIDESPIEVTDLVLLHQANDRLDKWAKIGVDRAKLPANWME
YGN TSAASIFILLSECVEQGLPLDGSQTVLLSGFGGLTWGTLILITI"
2611..2835
/gene="acc"
/notes="spr0378"
2611..2835
/gene="acc"
/codon_start=1
/transl_table=11
/label=spr0378
/product="Acyl carrier protein"
/protein_id="AAK99182.1"
/db_xref="GI:15457939"
/transl_table=11
/translation="MAVFEKVQETIIVEELGKGDASEVLTSTFDLDADSLDLFQVISE
IEDAFDICIENRDLKTVGDLVAVVEQAK"
2954..3928
/gene="fabK"

```

```

CDS
2954..3928
/gene="fabK"
/codon_start=1
/transl_table=11
/label=spr0379
/product="Enoyl-acyl carrier protein (ACP) reductase."
/protein_id="AAK99183.1"
/db_xref="GI:15457940"
/translation="MKRITBELLKIDYPIFOGMAWVADGLAGAVSAGGLGIIGGG
NAPKEVKANIDKISLTDKPKFKNIMLLSPFVEDIVLIEBGGVYVTTGAGNPSKY
MERFHEAGIIVIPVPSVALAKMEKIGADAVIAEGMEAGGHIGKLTMTTRVQVATA
VSIPIVIAAGIADGEGAAAGFMAEAVQVGTFRFVVAEKESNAHNPKEKILKARDIT
TISAQHFCHAVRAIKNQLTRDFELAEDKDFQEDPDLEIFPMGAGALAKAVVHGDE
GGSVMAGOIAGLVSKETAEELIKLDLYYGAACKIQEEASRWTVGVVRND"
3921..4841
/gene="fabB"
/notes="spr0380"
3921..4841
/gene="fabB"
/EC number="2.3.1.39"
/codon_start=1
/transl_table=11
/label=spr0380
/product="Malonyl acyl carrier protein transacylase"
/protein_id="AAK99184.1"
/db_xref="GI:15457941"
/transl_table=11
/translation="MTKTAFLPAGGAQYLGMRDFYDQYPIVKETIDRASQVLGYDL
RYLIDTEEDKLNQRYTQPAIATSAIVRYLLQKGYQPMVAGLSGEYSALVASGA
IDFEDAVALVAKRGNVHEEAPADSGKMAVLTVEVIEACAKASGLGVYTPANNY
TPAQIVIAQVAVDRAVELLOKAGAKLIFKVGSPHTSLEFPASOKLAETLAQVS
FSDFTCPVLVNGTEAAVMQKEDIAQLLTRQVKEPVRFYESIGVMQEAISNFIEIGPGK
VLSGFVKKIDQTAHLAHVEDQASLVALLEK"
4875..5606
/gene="fabG"
/notes="spr0381"
4875..5606
/gene="fabG"
/EC number="1.1.1.100"
/codon_start=1
/transl_table=11
/label=spr0381
/product="3-ketoacyl-acyl carrier protein reductase"
/protein_id="AAK99185.1"
/db_xref="GI:15457942"
/transl_table=11
/translation="MKLEHKNIFITGSSRGITGLAIHAKFAQAGANIVLNSGAISEEL
LAEFENYGIKVPISGDSVDFADAKRMIDQAIAGLSGVDVVLNNAGITQDTIMLKYTE
ADPEKVLKNTLTGAFNMVTQSVLKPMMKAREGAIINMSVWGLMGNIGQANVAASKAGL
IGFTKSVAREVASRNIRVNVIAPGMIESDMTAILSDKIKEATLAQIPMKETQAEQVA
DLTVFLAQDYLTGQVIAIDGGLSN"
5619..6863
/gene="fabF"
/notes="spr0382"
5619..6863
/gene="fabF"
/EC number="2.3.1.41"
/codon_start=1
/transl_table=11
/label=spr0382
/product="Beta ketoacyl-acyl carrier protein synthase"
/protein_id="AAK99186.1"
/db_xref="GI:15457943"
/transl_table=11
/translation="MKRMKLRNVVTVGYVTSPIGNTPEEPWNSLATKIGIGITKF
DHSDFDVHNAEIQDFPDQYFVKKDTNRFDNYSILYAAQEA/VNHNALDVEALNRD
RFGVIVASGIGIGIEIEDQVRLRHEKPKRVKPMTLPKALPNMAGNVAMPFGANGVC
KSNITACSSNDADGAPRSRFGQVLMVGGTEASITPFAIAGFQALTAETEDP
TRASIPFDKDRNGFVWGGSGMLVLESLHAEKRGATILAEVWGVTGNTCDYHMTSPH
PEGQAIKAIKLALEEAISPEQVAVYNAHCTSTPANEKSGSAIVAVLGRKEVPVST
KSFTGHLGLAAGAVEAIVTIEAMRHNFVPMTAGTSEVSDYIEANVVYQQGLEKIPYA
ISNTFGFGGHNAVLAFKWEKNR"
6866..7351
/gene="accB"
/notes="spr0383"

```

```

gene
CDS
2954..3928
/gene="fabK"

```

CDS

6866..7351
/gene="accB"
/BC_number="6.4.1.2"
/codon_start=1
/transl_table=11
/label="spr0383"
/product="Biotin carboxyl carrier protein of acetyl-CoA carboxylase"
/protein_id="AAK99187.1"
/db_xref="GI:15457944"
/translation="MNLDIKDLMTQFOSSLRERSYKNGTDELQFSKNEARPPVEVA
TQVAPAPLVAEPVAPVAPAEASVATGEGNVLVESPGLGVVYLAAG
DKPAFTVGVDSYKGGTLLIIEAMKVNIEIPAPKGVVTEILVSNEMVERGKGLVRI
K"
7348..7770
/gene="fabZ"

Alignment Scores:
Pred. No.: 5,79e-108 Length: 10925
Score: 1515.00 Matches: 305
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.67% Mismatches: 0
Query Match: 99.80% Indels: 0
DB: 1 Gaps: 0

US-09-308-397-2 (1-306) x AE008418 (1-10925)

Qy 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
Db 3921 ATGACTAAACAGCCCTTTTATTGCTGCTCAAGGTGCCAGTACTAGGATGGACGG 3980
Qy 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
Db 3981 GATTTCATGATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCCAGTCAGGTGCTA 4040
Qy 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60
Db 4041 GGTATGATTGGCTTATCTCATCGATCGGAGAGACAACTCAATCAGACCCGCTAT 4100
Qy 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGlnGluLysGly 80
Db 4101 ACGCAACCAACGATTCAGCAGCTTCGGTTGCTGCTATCTACCGTTTATTGCAAGAAAGGGC 4160
Qy 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
Db 4161 TATCAGCTCATATGTTGCTGCTTGTCTTTGGAGAAATCTTGCCTTGGTGCAAGC 4220
Qy 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyValaTyrMetGlu 120
Db 4221 GCGCCCTTGATTTGAACATCGGTTCGCTTGGTAGCTAAGCGTGGAGCCTATATGGA 4280
Qy 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
Db 4281 GAAGCGGCTCCCTGCTGACTCTGGCAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTC 4340
Qy 141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
Db 4341 ATTGAAGAGCCTGTCAAAGACCTTCTGAACTTGGAGTGGTACTCCAGCCCACTATAAC 4400
Qy 161 ThrProAlaGlnIleValIleAlaGlyValValAlaValAlaAspArgAlaValGluLeu 180
Db 4401 ACACCTGCACAAATCGTCAATCTCTGGAGAAGTGGTTCAGATTGATCGAGCGGTTGAACCT 4460
Qy 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr 200
Db 4461 TTGCAGAGACGAGGTGCCAAGCCTTGATTCTCTTAAGGTGTCAAGTCCCTTTCACACC 4520
Qy 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
Db 4521 TCTCTCTTGAACCTGCTAGCCAGAACTAGCTGAAACTCTGGCTCAGGTCAAGTTTTCAC 4580
Qy 221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle 240

Db 4581 GATTTCATTGTCCTCCATACAGAGCTGCTGTGATGCAAAAAGAGACATT 4640
Qy 241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
Db 4641 GCTCAGCTCTTGACCGCTCAGGTCAAGAACCCGTTCTGTTCTATGAAAGTATTGGGTC 4700
Qy 261 MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280
Db 4701 ATGCAGACGACGATATGCACTTTATCGAGATTGGACCGGGAAAGTCTTGTCAAGT 4760
Qy 281 PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
Db 4761 TTTGTTAAAAAATTGATCAAACTGCTACTAGCTCATGTGGAGATCAAGCAGATTTA 4820
Qy 301 ValAlaLeuLeuGluLys 306
Db 4821 GTAGCACTTTTAGAAAAA 4838

RESULT 8
AE006311 10701 bp DNA linear BCT 14-MAY-2001
LOCUS Lactococcus lactis subsp. lactis IL1403 section 73 of 218 of the
DEFINITION complete genome.
ACCESSION AE006311
VERSION AE006311.1
KEYWORDS GI:12723690
SOURCE Lactococcus lactis subsp. lactis.
ORGANISM Lactococcus lactis subsp. lactis.
REFERENCE 1 (bases 1 to 10701)
AUTHORS Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarme,K.,
Weissenbach,J., Ehrlich,S.D. and Sorokin,A.
TITLE The complete genome sequence of the lactic acid bacterium
Lactococcus lactis ssp. lactis IL1403
JOURNAL Genome Res. 11 (5), 731-753 (2001)
MEDLINE 21235186
PubMed 11337471
REFERENCE 2 (bases 1 to 10701)
AUTHORS Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarme,K.,
Weissenbach,J., Ehrlich,S.D. and Sorokin,A.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2001) INRA, Genetique Microbienne, Domaine de
Valvert, Jouy en Josas 78352, France
FEATURES
source
1..10701
/organism="Lactococcus lactis subsp. lactis"
/strain="IL1403"
/sub_species="lactis"
/db_xref="taxon:1360"
114..335
/gene="acpA"
/notes="L0183"
114..335
/gene="acpA"
/notes="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
PHOSPHOLIPID METABOLISM. General. KEY COMPONENT IN DE NOVO
FATTY ACID BIOSYNTHESIS"
/codon_start=1
/transl_table=11
/product="acyl carrier protein"
/protein_id="AAK04870.1"
/db_xref="GI:12723691"
181..1541
/translation="NAVPEKVDIIVDELGKEEVTLETSPELDADSLDFQIIND
IDPEFVEVDTEADMKTVADLVKYVNNK"
615..1541
/gene="fabD"
/notes="L0184"
615..1541
/gene="fabD"
/notes="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
PHOSPHOLIPID METABOLISM. General. FATTY ACID BIOSYNTHESIS"
/codon_start=1

Qy	21	AspPheTyrAspGlnTyrProIleValValysGluThrIleAspArgAlaSerGlnValLeu	40
Db	6950	GAATTTATGATAAATCTGCTATTGTAAGAAAAACCTTTGATCAAGCTAGTCAAGTATTG	6891
Qy	41	GlyTyrAspLeuArgTyrIleuLeuAspThrGluAspGlyLeuAsnGlnThrArgTyr	60
Db	6890	GGATAGCAATTCGCGCTTGGATGATAGTAGGAGCTTAAACTTAATCAAACTAGCAT	6831
Qy	61	ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGlyysGly	80
Db	6830	ACGCAACCAAGCTATTGATCATCATCAATGCTATTACCGTGTTTGGGCTTACATCAC	6771
Qy	81	TyrGlnProAspMetValAlaGlyLeuSerLeuGlyLeuTyrSerAlaLeuValAlaSer	100
Db	6770	GTTAAACCGGATATGATGCTGCTTCTCCTAGGAGATATCTACGTTTGTAGCATCA	6711
Qy	101	GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaValysArgGlyAlaTyrMetGlu	120
Db	6710	GGGCGACTCTCTTTGAAGATACCTATCTCTAGTAGCTAAGAGGCGCTTGATGGAG	6651
Qy	121	GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal	140
Db	6650	GAAGCAGCACCAAGAGATCTGGAAGATGGTTGCGGTTATGAATACAGACGTCGAAGTC	6591
Qy	141	IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn	160
Db	6590	ATCAGAGAGTCTGTCATATAGCTGCTAAACATGGAGTGGTTGCTCCAGCAACTATAT	6531
Qy	161	ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeu	180
Db	6530	ACTCCTTAGTCAAAATGTAATTTGGTGTGCAGACAGATGCTGTGAACGTGCGACTTGA	6471
Qy	181	LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr	200
Db	6470	TTAAGAAAGGGAGTTAAGCGTTTAATCCCTTTAAACGTGTCAGTCCCTTCCACACT	6411
Qy	201	AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer	220
Db	6410	GCCTTTGTAGAACCAAGTACGCGCTTGTGGCTAAAGAGTGTGAAAGATACAACTTCA	6351
Qy	221	AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetClnLysGluAspIle	240
Db	6350	GACTTCAAGATTCCTTTGGTTGGTAATACCGAGCTAATATTATGAAAAAGACCGTATC	6291
Qy	241	AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal	260
Db	6290	CCAGAACTATTACCGCTCAAGTCATGAGCGCTGTTCTGTTTATGACAGTGTGCGACT	6231
Qy	261	MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly	280
Db	6230	TTAGTAGAAGTGGCATAACCAATTCATTGAGGTAGGACCAAGGTAAGGTTTGTGACGT	6171
Qy	281	PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu	300
Db	6170	TTTGTGAGAAATTTGATAAAATTTACTATGCTAGTACTGTTTCAAAACATGCTAGCTTA	6111
Qy	301	ValAlaLeuLeuGluLys	306
Db	6110	CGTTATTTTATAGATAGA	6093
RESULT 10			
AE014165/c		50416 bp	DNA linear BCT 19-JUL-2002
LOCUS			
DEFINITION			Streptococcus pyogenes MGAS315, section 30 of 37 of the complete genome.
ACCESSION			AE014165 AE014074
VERSION			AE014165.1 GI:21905237
KEYWORDS			
SOURCE			Streptococcus pyogenes MGAS315.
ORGANISM			Streptococcus pyogenes MGAS315.
			Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
gene			
CDS			
gene			
CDS			
gene			
CDS			

1 (bases 1 to 50416)
 Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,
 Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
 Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
 Musser,J.M.
 Genome sequence of a serotype M3 strain of group A Streptococcus:
 Phage-encoded toxins, the high-virulence phenotype, and clone
 emergence
 Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)
 2 (bases 1 to 50416)
 Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,
 Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
 Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
 Musser,J.M.
 Direct Submission
 Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,
 Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,
 Hamilton, MT 59840, USA
 Location/Qualifiers
 1. 50416
 /organism="Streptococcus pyogenes MGAS315"
 /strain="MGAS315"
 /serotype="M3"
 /db xref="taxon:198466"
 /notes="group: A"
 117..1103
 /gene="SpyM3_1493"
 117..1103
 /gene="SpyM3_1493"
 /notes="best blastp hit: gb|AA198256.1| (AE010083) putative
 esterase [Streptococcus pyogenes MGAS8232]"
 /codon_start=1
 /transl_table=11
 /product="putative esterase"
 /protein_id="AA080100.1"
 /db_xref="GI:21905238"
 /translation="MKHPRKTLVTGLLTLCLPTPAAASSRSWSKSWIEQVFWKX
 DKSYKQDDPSQRYLDACREQSKPYQDTLVNGLVQENLYGMQVYVSMNDKRP
 DKTIYLAGSYVNDPTTYINMLKTLSTLDKIVLFTYKPAKPYTYNTMPKLVN
 LYQHYKYNQNVFLMGCSAGGLALGLAHNHSVPQKLVLLSPDLVDVMSHPEI
 KYEDADPLSSWGLKRVGLWAYSDNTNHIYVSPKNGPITYLPITLITGTREIFY
 PDJRDVAALKAAAHNITFTQSGMNHVPIYFIEBAKTAQYQIIDAINKTP"
 complement (1239..1589)
 /gene="rbfa"
 /notes="SpyM3_1494"
 complement (1239..1589)
 /gene="rbfa"
 /notes="best blastp hit: gb|AAK34465.1| (AE006601) putative
 ribosome binding factor A [Streptococcus pyogenes M1 GAS]"
 /codon_start=1
 /transl_table=11
 /product="putative ribosome binding factor A"
 /protein_id="AA080101.1"
 /db_xref="GI:21905239"
 /translation="MANHRIDRVGMBIKREVNILQKVRDPRVQGVTTTEVOMQGD
 SLAKVYVTMSDLASDNQKATQGLEKATGTIKRELKQLTWYKIPDLVFEKDNSIAYG
 NKIDQLRLDLNKS"
 complement (1755..4599)
 /gene="infB"
 /notes="SpyM3_1495"
 complement (1755..4599)
 /gene="infB"
 /notes="best blastp hit: gb|AAK34466.1| (AE006601) putative
 initiation factor 2 [Streptococcus pyogenes M1 GAS]"
 /codon_start=1
 /transl_table=11
 /product="putative initiation factor 2"
 /protein_id="AA080102.1"
 /db_xref="GI:21905240"
 /translation="MVEHAKYGLDVKSHASSVVEADAKKIISSFSKASKPDVTASQT
 VKPEVAQPSVTYVKSTGSHVEKTOVSKPSNFKAEREAREAKQAARQAQNGSSHR
 SQERGGYRQPNHHCNEQGDKXITHRSGDITNDKIEKASNVSRHNDHQLVGDEN
 RSFAKENHKNGRFTNOKKQGRQPSKPKIDFKARAAALKAEQNAEYRSQSETRFA

Db 30056 GTTAAACCGGATATGGTACTGGCTTCTTCTTAGAGATACTAGCTTTGGTAGCATCA 29997

Qy 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120

Db 29996 GGGGCACTCTCTTTTGAAGATACCTCTATCTTAGTAGCTTAAGAGAGCGCTTGATGGAG 29937

Qy 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140

Db 29936 GAGACGACCAACAGGATCTGGAAAGATGGTTGGCTTATGAATACAGAGCTGCAAGTC 29877

Qy 141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160

Db 29876 ATCGAAGAACTGTCTCAATAGCTGCTAAACATGAGTGCTGCTCCAGCAAACTATAAT 29817

Qy 161 ThrProAlaGlnIleValIleAlaGlyGluValAlaValAlaValAspArgAlaValGluLeu 180

Db 29816 ACTCTAGTCAAAATTGTAATTGGTGTGCAGACAGATGCTGTGAACGTGCAGTTGAATT 29757

Qy 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr 200

Db 29756 TTAAGGAAAGGGAGCTTAAGCGTTAATCCCTTTAAACGTGTGAGTCTTTCCACACT 29697

Qy 201 AlaLeuLeuGluProLysSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220

Db 29696 GCTTTTGTAGACCACTAGCCGCTGTGGCTAAGAGATTGGAAGATACAACTTCAGT 29637

Qy 221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle 240

Db 29636 GACTTCAAGATTCCTTTGTTGGTAAATACCGAAGCTAATATTATGGAAGAACCGGTATC 29577

Qy 241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260

Db 29576 CCAGAACTATTACCCGCTCAAGCTATGAGCCTGTTGTTTATGACAGTGTCCGACT 29517

Qy 261 MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280

Db 29516 TTAGTAGAAGTGGCATACAAATTCATTGAGTAGGACCGAGTAAGTTTGCACAGT 29457

Qy 281 PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300

Db 29456 TTGTGGAAGAAATTCATAAAATTTACTATGCTAGTGTGTAAGAACATGCTAAGCTTA 29397

Qy 301 ValAlaLeuLeuGluLys 306

Db 29396 CGTTATTTTATGATAGA 29379

AE010088 14456 bp DNA linear BCT 03-APR-2002

Streptococcus pyogenes strain MGAS8232, section 136 of 173 of the complete genome.

AE010088 AE009949

AE010088.1 GI:19748917

Streptococcus pyogenes MGAS8232.

Streptococcus pyogenes MGAS8232

Bacteria: Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.

1 (bases 1 to 14456)

Smoot, J.C., Barbican, K.D., Van Gompel, J.J., Smoot, L.M., Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Ricklefs, S.M., Porcella, S.F., Parkins, L.D., Beres, S.B., Campbell, D.S., Smith, T.M., Zhang, Q., Kapur, V., Daly, J.A., Veasy, L.G., and Musser, J.M.

Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks

Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)

21927593

11917108

2 (bases 1 to 14456)

Smoot, J.C., Barbican, K.D., Van Gompel, J.J., Smoot, L.M., Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Ricklefs, S.M., Porcella, S.F., Parkins, L.D., Beres, S.B., Campbell, D.S., Smith, T.M.,

Zhang, Q., Kapur, V., Daly, J.A., Veasy, L.G., and Musser, J.M.

Direct Submission

Submitted (31-JAN-2002) Laboratory of Human Bacterial Pathogenesis/Rocky Mountain Laboratories/NIH, 903 S. 4th St., Hamilton, MT 59840, USA

FEATURES

source

1..14456

/organism="Streptococcus pyogenes MGAS8232"

/strain="MGAS8232"

/db_xref="taxon:186103"

<1..101

/notes="phi_SDA"

misc_feature

86..101

/note="phi_SDA attr"

105..917

/gene="spyM18_1809"

105..917

/gene="spyM18_1809"

/notes="best blast match gb|AAK34481.1| (AE006602) conserved hypothetical protein [Streptococcus pyogenes M1 GAS]"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAL98329.1"

/db_xref="GI:19748918"

/translation="MTKLIADLDGTLHHDNTISTYTKTKIKAVQDKGHHVITSTG RPYRMALDYVLOLNKTPITTFNGALTMPEQKWAYEHNVILDKVGLRLLLKXQDDFG MFIASEKRYKVVITMTNPESIDPOLFGVDEITQDMALEITKTRPNALLMQTHED KYALAKMFAKPEDEIEMSGGPNLILSISKVNKAYALNVLLGIVNMDKDLIAF GDHNDTEMLAFAGTGTAKNNSPVLIPYADQOLNPSNEEDGVAKKLEELFL"

1269..2261

/gene="spyM18_1810"

1269..2261

/gene="spyM18_1810"

/notes="best blast match gb|AAK34482.1| (AE006602) mannose-specific phosphotransferase system component IIAB [Streptococcus pyogenes M1 GAS]"

/codon_start=1

/transl_table=11

/product="mannose-specific phosphotransferase system component IIAB"

/protein_id="AAL98330.1"

/db_xref="GI:19748919"

/translation="MGIGIIASHGKFAEGIHQSGSMIEGQEKVQVVTMPNEGPD LYCHFNNAIQCFDADDEILVLADLWSGSPFNQASRVAGENPDKMAITGLNDFMLIQ AYTERLMDAGAVEQVAANIIESKDGIKALPEDLNPEETAETAEKVNALQCAIPAG TVIGDGKILINLARVDTLLHGVQATANTFASKADRIIVASDEVAQDDLRKQLIKQAA PGSKVANVPISKLIEASKDPRFGNTHALILFQTPDALRAVEGVEINELNVGMAH STGKTWNVNLMSDKEDVATFEKLRDLGVTFDVRKVPNDSKKNLPFLIQKANIK"

2350..3159

/gene="spyM18_1811"

2350..3159

/gene="spyM18_1811"

/notes="best blast match gb|AAK34483.1| (AE006602) putative mannose-specific phosphotransferase system component IIC [Streptococcus pyogenes M1 GAS]"

/codon_start=1

/transl_table=11

/product="putative mannose-specific phosphotransferase system component IIC"

/protein_id="AAL98331.1"

/db_xref="GI:19748920"

/translation="MSDISIISAILVILVIAFFAGLEGILDQFMHQPVLVACTLIGLVLT GHLEAGVILGGTLQMLALGMANIGRAVADAALASVAAAIIMVSGDFTQKGTIRAYS TALPLAVAGLFTMTIVRTLSTALNHAGDKAAREGNFAGIERPHEFTALLOGLRIAPPA ALLVAFTSAVGSVLNAMENWLNNEGQIGGAMVAVGVAMVINMMATREWPEFALGF ALAATISQTLIANGVIGVIAFIYLNLSKKGNGNAGSADFIGDILEDY"

3176..4087

/gene="spyM18_1812"

3176..4087

/gene="spyM18_1812"

/note="best blast match gb|AAK34484.1| (AE006602)


```

putative mannose-specific phosphotransferase system
component IID [Streptococcus pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative mannose-specific phosphotransferase"
/system_component IID
/protein_id="AAL98332.1"
/db_xref="GI:19748921"
/translation="MTQIKLTKSDRQVWRVRSQFLOGSNVYERMQMGMWAYALIPAL
KLLTSPEDRAALREMEFNPYPVAAPLIGVTLAEEERANGPIDDKAQGVKI
GMGPLAGIDGDFVFWFTVRPILGALGASLASTGNVGPLLFFGWNLMRAFLLWYOE
FGYKAGEIKDKMSGGLQDITKGASILGMFILAIVQVRWVSINFIDILPGKQLSDGA
YVFPDGAAGAEIKTLTLANAIGMSLDKVAQTLOQLDLSLPLGLAGLLLTPLCWL
LKKVSPAIILIGLFAFGILAHLAGIM"
4201. 4560
/gene="spyM18_1813"
4201. 4560
/gene="spyM18_1813"
/note="best blast match gb|AAK34485.1| (AE006602)
conserved hypothetical protein [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL98333.1"
/db_xref="GI:19748922"
/translation="MAQSLNLSVEYKTKAVSYLGMGKGVGHILLGDKALEFYNDKXNV
DYIOIPTWAINHIGANVSRRKVSRRHPEIFDQGFPLFASGDSGKILKIRQIHGNEKV
ITLPLMQTRINKFFRH"
4952. 6229
/gene="serS"
/note="spyM18_1814"
4952. 6229
/gene="serS"
/note="best blast match gb|AAK34486.1| (AE006602)
putative seryl-tRNA synthetase [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="putative seryl-tRNA synthetase"
/protein_id="AAL98334.1"
/db_xref="GI:19748923"
/translation="MLDLKIRITDFDTVAALKNRGVSEDTLHLKELDEKRTLLAVQ
SEELKASENTASAAIAKQKQEDATQIADMOKVSADIKINDQLVALDQVADIIIT
VLENPDSVPFGADEEDNEIRRWGTPRODFEVKAHMDLSDLDLIDLWDERAKVTG
ARFYKNGLARGALRYNFMLEHKEGYOEIITPMVNHDSMFQGGYKPKEDTF
ELADTNFLVLPTEAVPTNYRGEILDKELPIYFTAMSPSRSEAGSAGRDTRGLIR
LHQPHKVMVKFAPEESYQLEKMTANAENILQKLGLPYRVSISLCTGDMGFGAAKTY
DLEVIIPAONTYREISSCNTEDFOARRAQIRYRDEADGKVKLLHTLNGSLAVGRTV
AAILENYQNEGSGVTIPEVLRYPMGGETVISPK"
complement(6462. 7232)
/gene="acca"
/note="spyM18_1815"
complement(6462. 7232)
/gene="acca"
/note="best blast match gb|AAK34487.1| (AE006603)
putative acetyl-CoA carboxylase alpha subunit
[Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative acetyl-CoA carboxylase alpha subunit"
/protein_id="AAL98335.1"
/db_xref="GI:19748924"
/translation="MTDVSVLKEARDQGBLTLDVYANLIPDDFMELHGDHPSDDGA
IVGFLAYLAQPVTVIGIQKNLOHNLARNFQCPNPEGYRKRLRMKQKRFKRPVW
FINTAGYFGVGAEEGQGEATKKNLMNSDLKVPILAIIGEGSGSGALALAVADQ
VMUENTMYAVLSEGSATILWKDGSRAEALMKITAGELYQMGIVDRIPIPEHGYF
SSEIVDIIKANLIPDSLOAKPLDQLLDERYQRFKY"
complement(7229. 8095)
/gene="accD"
/note="spyM18_1816"
complement(7229. 8095)
/gene="accD"

```

```

/note="best blast match gb|AAK34488.1| (AE006603)
putative acetyl-CoA carboxylase beta subunit
[Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative acetyl-CoA carboxylase beta subunit"
/protein_id="AAL98336.1"
/db_xref="GI:19748925"
/translation="MALFRKDKVIRITPNPNFLKGSVSHNVPEVDELPACPKACKHM
IYKDLGLAKICPTCSYNFRISAQERLTLDVDEGSFQELTSIEDPLRFPYQEKL
QKAKETTLGHAIVTKGNVRGQOIALIMDSHFIMASMGTVGVEKIRLPELAIEEN
LPVFTTASGARQOEGIMSQMAKVSAAVKRHSNAGLYLTLDITDTTGGVTAFA
MEGDIILAEPOSQVFGACRRVETTRNRELPDDFKAEFLDQHGFDVAIVKRTURDK
TAHLVAPRHGGGQ"
complement(8104. 9468)
/gene="accC"
/note="spyM18_1817"
complement(8104. 9468)
/gene="accC"
/note="best blast match gb|AAK34489.1| (AE006603)
putative acetyl-CoA carboxylase biotin carboxylase subunit
[Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative acetyl-CoA carboxylase biotin
carboxylase subunit"
/protein_id="AAL98337.1"
/db_xref="GI:19748926"
/translation="MFKXILLANRGEAVRIIRAARELGISTVAVYSEADKEALHTIL
Alignment Scores:
Pred. No.: 4,36e-70 Length: 14456
Score: 1030.00 Matches: 208
Percent Similarity: 80.72% Conservative: 39
Best Local Similarity: 67.97% Mismatches: 59
Query Match: 67.85% Indels: 0
DB: 1 Gaps: 0
US-09-308-397-2 (1-306) x AE010088 (1-14456)
QY 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
13330 ATGACAAAACAGCTTTTATTTCCTCGGTCAAGTGCGCCAAAATTAGGATGCCAAGG 13271
QY 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
13270 GATTTTATGATACTTTTCTGCTATTGTAAGAAAACTTTTGTATCAAGCTAGTCACTATG 13211
QY 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60
13210 GGATACGATTTCGCCCGTTTGNATTGATGACGAGTTAAACCTTAATCAACTAGCTAT 13151
QY 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGlnGlnLysGly 80
13150 ACCCAACAGCTATTTCGACATCATCAATTCGTTATTACCGTGTGGGCTTACATCAC 13091
QY 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
13090 GTTAAACCGGATGATGCTAGCTGGTCTTTCCCTTAGGAGAACTACTCAGCTTTGGTGCATCA 13031
QY 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyValaTyrMetGlu 120
13030 GGGGCACCTCTCTTTTGAAGATACCCTATCCTTAGTAGCTAAGAGAGCGCGCTTGATGGAG 12971
QY 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
12970 GAAGCAGCACCACAAAGATCTGGAAGATGGTTCGCCGTATTGAATACAGAGCTCAAGTC 12911
QY 141 IleGluGluAlaCysGlnLysAlaSerGlnLeuGlyValValThrProAlaAsnTyrAsn 160
12910 ATCGAAGAGTCTGTCAATAGCTGCTAAACATGAGTGGTGTGCTCCAGCAACTATAT 12851
QY 161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeu 180

```

```
Db 12850 GCTCCTAGTCAAAATTGTAATTGGTGTGACACAGATGCTGTGAACCTGCGAGTTGAACCT 12791
Qy 181 LeuGlnGluAlaGlyAlaTysArgLeuIleProLeuLysValSerGlyProPheHisThr 200
Db 12790 TTAAGGAAAGGGAGTAAAGCGTTTAAATCCCTTTAAACGCTGTGACGCTCTTCCACACT 12731
Qy 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
Db 12730 GCCTTGTGTAAGACAGCTAGCCCTGTGGCTTAAAGAGTGTGAAGATACAACTTCACT 12671
Qy 221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle 240
Db 12670 GACTTCAAGATTCTTGGTGTGTAATACCGAGCTTAATATATGAAAAAGACCGTATC 12611
Qy 241 AlaGlnLeuThrArgGlnValLysGluProValArgPheThrGluSerIleGlyVal 260
Db 12610 CCGAAGACTATACCCCGTCAAGTCATGGAGCCCTGTGCTTTATGACAGTGTGCGACT 12551
Qy 261 MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280
Db 12550 TTAGTAGAAGTGGCATACACAACTTCAATGAGTAGGACCGGTAAAGTTTGCAGGT 12491
Qy 281 PheValLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
Db 12490 TTTGTGAAGAAATTTGATAAAATTTACTATGTACTGTGTTGCAAAACATGCTAAAGCTTA 12431
Qy 301 ValAlaLeuLeuGluLys 306
Db 12430 CGTTTATTTTAGATAGA 12413

RESULT 12
LOCUS AR193756 1196 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 8 from patent US 6348328.
ACCESSION AR193756
VERSION AR193756.1 GI:20240348
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1196)
AUTHORS Black, M. Terence., Hodgson, J. Edward., Knowles, D. Justin, Charles.,
Nicholas, R. Oakley., and Stodola, R. King.
TITLE Compounds
JOURNAL Patent: US 6348328-A 8 19-FEB-2002;
FEATURES
    Location/Qualifiers
    source 1..1196
    /organism="unknown"
BASE COUNT 355 a 223 c 286 g 332 t
ORIGIN

Alignment Scores:
Pred. No.: 4,33e-58 Length: 1196
Score: 859.00 Matches: 180
Percent Similarity: 98.37% Conservative: 1
Best Local Similarity: 97.83% Mismatches: 1
Query Match: 56.52% Indels: 2
DB: Gaps: 0

US-09-308-397-2 (1-306) x AR193756 (1-1196)

Qy 125 AlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGluGluAla 144
Db 2 GCTGAAGTGGCAGATGTTAGCAGTTCTCAATACCCAGTAGAGTCAATGAAAGGCC 61
Qy 145 CysGlnLysAlaSer-GluLeuGlyValValThrProAlaAsnTyrAsnThrProAlaGln 164
Db 62 TGTCAAAAGCTTCTGGAAGTGGAGTGGTACTCCAGCCCACTATAACACACCTGCACA 121
Qy 164 nIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeuGlnGluAl 184
Db 122 AATCGTCATTGCTGGAGAGTGGTTCAGTTGATCGAGCGGTGAACCTTTTCAAGAGC 181
```

```
Qy 184 aGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeuLeuGln 204
Db 182 AGGTGCCAAAACGCTTGTATCTCTTAAGGTGTGAGTCCCTTTTACACACCGCTCTCCCTTGA 241
Qy 204 uProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPheThrCys 224
Db 242 GCCTGTAGCCAGAACTAGCTGAACCTAGCTCAGTAAAGTCTTTTTCAGATTTCCTTG 301
Qy 224 sProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIleAlaGlnLeuLe 244
Db 302 TCCCTAGTCCGCAATACAGAAGCTGTGTGATGCAAAAAGAGACACATTGCTCAGCTCTT 361
Qy 244 uThrArgGlnValLysGluProValArgPheThrGluSerIleGlyValMetGlnGluAl 264
Db 362 GACCGCTCAGTCAAGGACCCGCTGCTTTCTATGAAGATTTGGGGTCTATGAAGAAGC 421
Qy 264 aGlyIleSerAsnPheIle-GluIleGlyProGlyLysValLeuSerGlyPheValLysL 284
Db 422 AGCATTAAGCAACTTTATTCGAGATTGGACCGGGAAGTCTTGTGAGGTTTGTAAAA 481
Qy 284 yIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeuValAlaLeuL 304
Db 482 AAATGTGATCAAACTGCTCACTTAGCTCATGTGAAGATCAAGCGAGTTTAGTAGCACTTT 541
Qy 304 euGluLys 306
Db 542 TAGAANA 549

RESULT 13
LOCUS AE013105/c 11845 bp DNA linear BCT 09-MAY-2002
DEFINITION Thermoanaerobacter tengcongensis strain WB4T, section 132 of 244 of
the complete genome.
ACCESSION AE013105 AE008691
VERSION AE013105.1 GI:20516482
KEYWORDS
SOURCE Thermoanaerobacter tengcongensis.
ORGANISM Thermoanaerobacter tengcongensis
REFERENCE 1 (bases 1 to 11845)
AUTHORS Bao, Q., Tian, Y., Li, W., Xu, Z., Xuan, Z., Hu, S., Dong, W., Yang, J.,
Chen, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y., Ling, L.,
Tan, H., Chen, R., Wang, J., Yu, J., and Yang, H.
TITLE A Complete Sequence of the T. tengcongensis Genome
JOURNAL Genome Res. 12 (5), 689-700 (2002)
FEATURES
    MEDLINE 21992816
    PUBMED 11997336
    REFERENCE 2 (bases 1 to 11845)
    Bao, Q., Xu, Z., Hu, S., Dong, W., Chen, Y., Wang, J., Yu, J. and Yang, H.
    Direct Submission
    Submitted (07-OCT-2001) Beijing Genomics Institute/Genomics and
    Bioinformatics Center, Institute of Genetics and Development,
    Chinese Academy of Sciences, Beijing Airport Industrial Zone B6,
    Beijing 101300, China
    3 (bases 1 to 11845)
    Li, W., Xuan, Z., Yang, J., Ling, L. and Chen, R.
    Direct Submission
    Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of
    Sciences, Beijing 100101, China
    4 (bases 1 to 11845)
    Tian, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y. and Tan, H.
    Direct Submission
    Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy
    of Sciences, Beijing 100080, China
    Location/Qualifiers
    1..11845
    /organism="Thermoanaerobacter tengcongensis"
    /strain="WB4T"
    /db_xref="taxon:119072"
    complement(76..1153)
    /gene="EUP3"
    source
    gene
```

```

CDS
/note="TTE1468"
/complement(76. .1140)
/gene="ELP3"
/note="Best Blastp hit = gi|7462101|pir|A72295 conserved
hypothetical protein - Thermotoga maritima (strain MSB8)
gi|4981648|gb|AA036179.1|AE001769.9 (AE001769) conserved
hypothetical protein [Thermotoga maritima], score 202,
E-value 6.00E-51"
/codon_start=1
/transl_table=11
/product="ELP3 component of the RNA polymerase II complex,
consists of an N-terminal BioB/LipA-like domain and a
C-terminal histone acetyltransferase domain (only
BioB/LipA-like domain in Thermotoga)"
/protein_id="AAM24690.1"
/db_xref="GI:20516483"
/translacion="MKRKNYIIIFIPHWGCSFKVCNQNLTGKKQVTVVEEVET
TESLTKIPGSEVSEVFGSGFTGPIEKQNDYLSIAREYLRKGOIGIRLSTRDPY
ITEILNLKKHDSVIELGVQSMEDDLKSRGHTSDEVVNAVLRIRYDFKLGLO
LMIGLPGDTBEKSLNTAKIVSLKPDVRYTPTLVIRNTYLEKYKEGYTFLSLQLO
INISKWYIIFVKGIDIVRIGLQTDINFNKDVVAGPFPAMGOLVSSIMLDVLK
EIVKEGIGKKISIFCNERKLSVIGQKYNKRFLEDKYGLOVQVKLEKIEEDKIV
LSYDKLYPSITDFKDNF"
/complement(1091. .1108)
/terminator
/gene="ELP3"
/note="putative rho-independent transcription terminator"
/complement(1137. .1837)
/gene="Rnc"
/note="TTE1469"
/complement(1137. .1823)
/gene="Rnc"
/note="Best Blastp hit = gi|10175109|dbj|BAB6208.1|
(AP001515) ribonuclease III [Bacillus halodurans], score
193, E-value 1.00E-48"
/codon_start=1
/transl_table=11
/product="dsRNA-specific ribonuclease"
/protein_id="AAM24691.1"
/db_xref="GI:20516484"
/translacion="MLTYLEQKINYEKKFXTLLLEALTHSSWAHEGKNEKVSNERLEF
LGDVSLVISEVLYKNRKLERSLSKRAEIVCPSLARCAKIELGSGFRMKGGE
EISGCRDRDGIADAMALLAAVLDGLLEAVRRVLDLFKEIIVDLKGIIVRDYKT
RLQBVQSMVSGKITYLEVVEIGDPDHNTFTVQVKGIDVVLGIGQGSKESEQAAM
EALSKLGLK"
/complement(1149. .1153)
/RBS
/gene="ELP3"
/complement(1152. .1352)
/gene="Rnc"
/note="Pfam match to entry dsrm, Double-stranded RNA
binding motif, score 85.2, E-value 1.30E-21"
/misc_feature
/complement(1134. .1706)
/gene="Rnc"
/note="Pfam match to entry Ribonuclease_3, RNase3 domain.,
score 139, E-value 8.40E-38"
/RBS
/complement(1833. .1837)
/gene="Rnc"
/complement(1849. .3092)
/gene="FabB"
/note="TTE1470"
/complement(1849. .3078)
/gene="FabB"
/note="Best Blastp hit = gi|7433750|pir|G69842
3-oxoacyl-[acyl-carrier-protein] synthase (EC 2.3.1.41) -
Bacillus subtilis gi|2633471|emb|CAB12975.1| (Z99109)
similar to 3-oxoacyl- acyl-carrier protein synthase
[Bacillus subtilis] gi|2633488|emb|CAB12991.1| (Z99110)
similar to 3-oxoacyl- acyl-carrier protein synthase
[Bacillus subtilis], score 496, E-value 1.00E-139"
/codon_start=1
/transl_table=11
/product="3-oxoacyl-(acyl-carrier-protein) synthase"
/protein_id="AAM24692.1"
/db_xref="GI:20516485"

```

```

/misc_feature
/translacion="MNRVVTGLGVISPLGNTVEKFWNSLINGENGLVTKFDPSF
PTKYAEVKDPDFLYIDKKARRMDRFQFAVASTKLALEDAGIELGLVDLNRGVI
YGTGIGTETENQMYEKSPKSPFFIPMMIANWAGOIATIFGLKGINETIVN
ACASTNAIGFAKAIQRGDLIVTGTETAITPMSFAGCAMKSTNPDPSASR
PFDLNRDGFVNGEGSATILLESHAQKRGAIVAEIIGYGATDDAYHITAPAGEG
AARAMEALKDQKSYDMVDVYNAGHTSTENDKFTETWAIKVKFKDHYKLCVSNKS
MLGHLGAAGAWEAVATVLTITHGIVPPTINYQTPDPCDLNYPNKAIEREVNVAIS
NSFGPGHNATLLEKFF"
/complement(1855. .2325)
/gene="FabB"
/note="Pfam match to entry ketoacyl-synt_C, Beta-ketoacyl
synthase, C-terminal domain, score 238.8, E-value
7.70E-68"
/complement(2347. .3078)
/gene="FabB"
/note="Pfam match to entry ketoacyl-synt, Beta-ketoacyl
synthase, N-terminal domain, score 295.3, E-value
7.70E-85"
/RBS
/complement(3088. .3092)
/gene="FabB"
/complement(3111. .3127)
/note="putative rho-independent transcription terminator"
/complement(3133. .3432)
/gene="Acpp"
/note="TTE1471"
/complement(3133. .3432)
/gene="Acpp"
/note="Best Blastp hit = gi|3912967|sp|O6761.1|ACP_AQUAE
ACYL CARRIER PROTEIN (ACP) gi|7442262|pir|A70448 acyl
carrier protein - Aquifex aeolicus
gi|3984024|gb|AAC07567.1| (AE000752) acyl carrier protein
[Aquifex aeolicus], score 75.5, E-value 2.00E-13"
/codon_start=1
/transl_table=11
/product="Acyl carrier protein"
/protein_id="AAM24693.1"
/db_xref="GI:20516486"
/translacion="MLTTEELNCNIIISVQFLRGCKGMIFEKVNITAEQIGIDPEEI
TMESFIDLGADSLDIVELIMALEEEDIEIPDEAKIKTVGDDVEYLSNIVE"
/complement(3151. .3354)
/gene="Acpp"
/note="Pfam match to entry pp-binding, Phosphopantetheine
attachment site, score 95.2, E-value 9.80E-25"
/complement(3407. .4163)
/gene="FabG3"
/note="TTE1472"
/complement(3407. .4150)
/gene="FabG3"
/note="Best Blastp hit = gi|1502421|gb|AAC44307.1|
(U59433) 3-ketoacyl-acyl carrier protein reductase
[Bacillus subtilis], score 265, E-value 6.00E-70"
/codon_start=1
/transl_table=11
/product="Dehydrogenases with different specificities
(related to short-chain alcohol dehydrogenases)"
/protein_id="AAM24694.1"
/db_xref="GI:20516487"
/translacion="MVDEKRVAFVTGSGRIGRAIVRLAKDGNIAITYVKDKSAE
EVVEVKHGVDALAKVDKSVKEKAVEKVEEFGSDVVVNNAGITKDNLLIKM
EEEDQVIDVNLKGFNIVFKASKMIKKRKGKIIINISVVLGMVNGVQYAAASKA
GIIGTKSAVELASRGITVNAVAPGFIETDMTNVLKEDIKEAMLKSIPLKRGKPEE
VAEVVAFASSASDYITGOVINVDGVMV"
/complement(3428. .3520)
/gene="FabG3"
/note="Pfam match to entry adh_short_C2, Short chain
dehydrogenase/reductase C-terminus, score 59.2, E-value
3.40E-15"
/complement(3578. .4135)
/gene="FabG3"
/note="Pfam match to entry adh short, short chain
dehydrogenase, score 283.1, E-value 3.60E-81"
/complement(4154. .5093)
/gene="FabD"

```

```

CDS
/note="TFE1473"
complement(4154..5083)
/genes="FabD"
MALONYL COA-ACYL CARRIER PROTEIN TRANSACYLASE (MCT)
gi|7433733|pir|H69620|acyl-carrier-protein|
S-malonyltransferase [EC 2.3.1.39] fabD [validated] -
Bacillus subtilis gi|2337819|emb|CAA74249.1| (Y13937)
putative fabD protein [Bacillus subtilis]
gi|2633962|emb|CAB13463.1| (Z99112) malonyl CoA-acyl
carrier protein transacylase [Bacillus subtilis], score
276, E-value 2.00E-73"

Alignment Scores:
Pred. No.: 2,11e-50 Length: 11845
Score: 775.50 Matches: 154
Percent Similarity: 69.97% Conservative: 58
Best Local Similarity: 50.83% Mismatches: 90
Query Match: 51.09% Indels: 1
DB: 1 Gaps: 1

US-09-308-397-2 (1-306) x AE013105 (1-11845)

Qy 3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22
Db 5080 AAAATAGCTTTATATACCGCGCCCAAGGTGCCAGTACGAGGATGGGAAAGGAATA 5021
Qy 23 TyrAspGlnTyrProIleValIysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
Db 5020 TACGAAATACGAAAGCGGAAAGAAATTTTGAAAGCAGATGAGCTTTAGCTTTT 4961
Qy 43 AspLeuArgTyrLeu---IleAspThrGluGluAspLysLeuAsnGlnThrArgTyrThr 61
Db 4960 AACATATCTAACTGTCTTGAAGGGCCTGAAGAGAGCTTATGAAGCGGAGAACACA 4901
Qy 62 GlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGlyTyr 81
Db 4900 CAGCCGGCATTTTGACACTTTCAGTGGCTTTCAGCAAGGGTTTTCGAAAGAGAGTGA 4841
Qy 82 GlnProAspMetValAlaGlyLeuSerLeuGlyTyrSerAlaLeuValAlaSerGly 101
Db 4840 AAACCTGATGTACACACAGGTTTAAGTTTAGGAGAGTATTCCTCTTGACTGCAGAA 4781
Qy 102 AlaLeuAspPheGlnAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 121
Db 4780 GCTTTAGATTTTGAGATCCAGTACGACTGTGCAAAAGAGAGGAAAGTACATGCAGAA 4721
Qy 122 AlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIle 141
Db 4720 GTGTCCTCCAGAGGAGTCCGCACATGGCAGCAATCTTGGACTTCCCAATCAGGAAGTT 4661
Qy 142 GluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThr 161
Db 4660 GAGGAATAATGCCGATTCATCAGAGTCCGCTGTGTGAGAACCTGCTAACTACAACTGT 4601
Qy 162 ProAlaGlnIleValIleAlaGlyGluValAlaValAlaAspArgAlaValGluLeuLeu 181
Db 4600 CCGGTCATATAGTCTCTCGGGAGAGTTAGGCTGTGAGAGAGCTGTGGAACTACGC 4541
Qy 182 GlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAla 201
Db 4540 AAGGAGAGAGGAGCTAAGAAAGCAGTAGTGTCTGCAGTAAGTGCCTTCCCTCAGTCAGC 4481
Qy 202 LeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAsp 221
Db 4480 ATGTTGAAAGAGCTGGAGACTTCTAGCAAGAGATTTGGATAAAGTACAGATAAAGGAC 4421
Qy 222 PheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIleAla 241
Db 4420 TTCAAGGTTCCCTGTGATACAAATGTTACTGCGGATTAAGTCCAAAGAGATAAGGTTAA 4361
Qy 242 GlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMet 261
Db 4361 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT

```

```

Db 4360 GAATTACTCTATAAAACAGGTGAGTCTCCCGTCTCTGGAGCAGTCTGTGAGAAAATG 4301
Qy 262 GlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGlyPhe 281
Db 4300 ATAGAAGCGGTGGATACCTTTTCATAGAAATAGGACCTGGAAGACTTTGAGCGGATTT 4241
Qy 282 ValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeuVal 301
Db 4240 GTCAAGAAATTCATAGACGAGCAGTGTGTTGAATTTTGGATGAGGAGTCTCTCATG 4181
Qy 302 AlaLeuLeu 304
Db 4180 AAGGCTCTC 4172

RESULT 14
AE007854/c
LOCUS
DEFINITION
  Clostridium acetobutylicum ATCC824 section 342 of 356 of the
  complete genome.
ACCESSION
  AE007854 AE001437
VERSION
  AE007854.1 GI:15026666
KEYWORDS
  Clostridium acetobutylicum.
SOURCE
  Clostridium acetobutylicum.
ORGANISM
  Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
  Clostridium.
REFERENCE
  1 (bases 1 to 13165)
  Nolling,J., Breston,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
  Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
  Tatusov,R.L., Sabatche,F., Doucette-Stamm,L., Soucaille,P.,
  Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
  Genome sequence and comparative analysis of the solvent-producing
  bacterium Clostridium acetobutylicum
  J. Bacteriol. 183 (16), 4823-4838 (2001)
JOURNAL
  21359325
MEDLINE
  11466286
PUBMED
REFERENCE
  2 (bases 1 to 13165)
  Childress,D., Zeng,Q. and Smith,D.R.
  Direct Submission
  Submitted (24-JUL-2001) GTC Sequencing Center Production,
  Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
  Beaver Street, Waltham, MA 02453-8443, USA
FEATURES
  Location/Qualifiers
    1..13165
       /organism="Clostridium acetobutylicum"
       /strain="ATCC 824"
       /db_xref="ATCC:824"
       /db_xref="taxon:1488"
       complement(74..499)
       /genes="CAC3571"
       /notes="fabZ"
       complement(74..499)
       /genes="CAC3571"
       /codon_start=1
       /transl_table=11
       /product="Hydroxymyristoyl- (acyl carrier protein)
       dehydratase"
       /protein_id="AAK81494.1"
       /db_xref="GI:15026667"
       /translation="MSLSIEQIMEIIPHPMLVDVREIEIFGKRAVGYNVTFNEQ
       IFQHYEPKPIPMGVLMIEBALQGVAILSLDKYKKKPIILGAVXNAKFRMVVPGD
       VLKLEIIVKVGPGAGIKGIATVNGEKAVERAIEITFMIV"
       complement(520..999)
       /genes="CAC3572"
       /notes="accB"
       complement(520..999)
       /genes="CAC3572"
       /codon_start=1
       /transl_table=11
       /product="Biotin carboxyl carrier protein of acetyl-CoA
       carboxylase"
       /protein_id="AAK81495.1"
       /db_xref="GI:15026668"

```


Score: 710.50 Matches: 146
Percent Similarity: 67.11% Conservative: 56
Best Local Similarity: 48.50% Mismatches: 96
Query Match: 46.81% Indels: 3
DB: 1 Gaps: 3

US-09-308-397-2 (1-306) x AE007854 (1-13165)

QY 3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyWetGlyArgasphe 22
Db 3980 AAGATTGCATTTGTTTTCAGACAGGGTTCGCAGTATGTTGGAATGGGAAGATCTA 3921

QY 23 TyrAspGlnTyrProIleValIysGluThrIleAspArcAlaSerGlnValLeuGlyTyr 42
Db 3920 TACGATAATTACCAAGCCCAAGGAACCTTTTGTAAAGCTGATGAGTGGTTCGCTTT 3861

QY 43 AspLeuArgTyrLeu---IleAspThrGluGluAspLysLeuAsnGlnThrArgTyrThr 61
Db 3860 AAAATCAGTGAATATGCTTCGAAGGAAGATGAGAAATTAATTTAACTGAGAACAT 3801

QY 62 GlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeu---GlnGluLysGly 80
Db 3800 CAGCCAGAGTTTAACTACTAGTATGCTTAAAGAGCACTTGAAGAAGAAAAGGA 3741

QY 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
Db 3740 ATAAACCTCATGTTCTGCAGCTTAAGCTTAGGAGATGATTCGCACATGTTGCACT 3681

QY 101 GlyValaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
Db 3680 GGAAGTTTTCATTTGAGATGCGAGTAAGATTAGTGAAAAAAGAGGAGATATATGCA 3621

QY 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
Db 3620 GAAGCAGTACCAAAAGGAATAGTACTATGCTGCTATTAATCGGACTTGAAGTGTAT 3561

QY 141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaSerTyrAsn 160
Db 3560 GTTAGAGGTATATGCCCTGAGCAATCAAAGGAAGTAAGTAGAGTAGCAATTAATAT 3501

QY 161 ThrProAlaGlnIleValIleAlaGlyGluValAlaValAlaAspArgAlaValGluLeu 180
Db 3500 TGTCAGGACAAATTCGTATAGCAGGAGAAGTTAAAGCTGTTGAAAGCCCTGTAAG 3441

QY 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr 200
Db 3440 TTAAGAAAGCGGAGCAAGCAAGACAGTAATGCTTTCAGTAAGTGGACCATTTCCACA 3381

QY 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
Db 3380 TCATGCTAAAGAGTCTGCTGAAAATTTAGAAGAAGATTTAAAGATATAATATAGAA 3321

QY 221 AspPheThrCysProLeuValGlyAsn---ThrGluAlaAlaValMetGlnLysGluAsp 239
Db 3320 GATATGAAGTACCACTTATCAAAATGTAACAGGAGACTATGTAGAAGATAAGATCAG 3261

QY 240 IleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGly 259
Db 3260 ATAAAGGTTTATTGAAAAGCAAGTCAGCTCTGTTAGATGGGAAGATACATATAGA 3201

QY 260 ValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSer 279
Db 3200 AGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3141

QY 280 GlyPheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSer 299
Db 3140 TCATTTATAAGAAATTAATAGAAAAAATGACTATATTTAATTTGAAAAAGCTGAAGAT 3081

QY 300 Leu 300
Db 3080 TTA 3078

RESULT 15

AP003189 298050 bp DNA linear BCT 10-JUL-2002
LOCUS Clostridium perfringens str. 13 DNA, complete genome, section 5/10.
DEFINITION AP003189 BA000016
ACCESSION AP003189.2 GI:18146727
VERSION
KEYWORDS
SOURCE
ORGANISM
Clostridium perfringens str. 13 (strain:13) DNA.
Clostridium perfringens str. 13
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE
1 Shimizu,T., Ohtani,K., Hirakawa,H., Ohshima,K., Yamashita,A.,
Shiba,T., Ogasawara,N., Hattori,M., Kuhara,S. and Hayashi,H.
Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater
Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)
2 21664373
PUBMED 11792842
REFERENCE 2 (bases 1 to 298050)
AUTHORS Shimizu,T.
DIRECT SUBMISSION
SUBMITTED (15-FEB-2001) Tohru Shimizu, Institute of Basic Medical
Sciences, University of Tokuba, Department of Microbiology; 1-1-1
Tennohdai, Tsukuba, Ibaraki 305-8575, Japan
(E-mail:tshimizu@md.tsukuba.ac.jp, Tel:81-298-53-3354,
Fax:81-298-53-3354)
COMMENT On Jan 14, 2002 this sequence version replaced gi:18144662.
FEATURES
Location/Qualifiers
1..298050
/organism="Clostridium perfringens str. 13"
/strain="13"
/db_xref="taxon:195102"
/notes="anaerobic pathogen for gas gangrene"
complement (34..1056)
/genes="CPE1002"
complement (34..1056)
/genes="CPE1002"
/notes="340 aa, no significant homology."
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB80708.1"
/db_xref="GI:18144663"
/translation="MSKNKINFKMNSDKVLINSLVLDGSLKSLKIDIOKKD
IIFPMEEFFSTGYPKKKEESTSIDIKPSVETSKSSLPKTEREDVQSIIDSI
KKKINRDEITNILLPEIKTSKVKEDNEDNTKTSQMTLEIPVSVTRARRK
KKKTLIKMISYFKSKIFLIKKNKSSKIECNDFNKVKKVINKNSTILLEE
VAIEKLVNDKIETTSHGNSKLPFEIKPKPKRRRIKPKVDCDLDEINFLND
YSLEDLKKPFDLKLNSDTIKNINSDLSLSNRDKSLERVCNLEKKEINELEFFLEIF
QAMSNK"
1361..1540
/genes="CPE1003"
1361..1540
/genes="CPE1003"
/notes="59 aa, similar to gnu:AP001515_87 BH2353 gene
product from Bacillus halodurans (79 aa); 46.7% identity
in 45 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB80709.1"
/db_xref="GI:18144664"
/translation="MNIDELTKRINELHKKHKEEGLSDEHKEERELRKEYINRFSKN
LRQLKGIPKNNKN"
1552..2475
/genes="hprk"
1552..2475
/genes="hprk"
/notes="307 aa, similar to gnu:AP001519_105 Hpr (Ser/Thr)
protein kinase/phosphatase from Bacillus halodurans (310
aa); 45% identity in 282 aa overlap
CPE1004"
/codon_start=1

```

/transl_table=11
/product="probable HPr Ser/Thr protein kinase/phosphatase"
/protein_id="BAB80710.1"
/db_xref="GI:18144665"
/translation="MGVTEKLIKDFLEVIQIGENVPINVDNVRPGQLAGFYNY
FAPRIOVIGAEBSFLEDSPLDKRLANKRFSFDISCLITRGLEIHEELKAARK
RNLWLSDMVTKFEISKIMVYLSDKWAPETRLHGLVDVYVGIMLITGESGIGKSET
ALELIKGRHLVTDADVDKIDGDLIGRFEITFGMLEVRGMIIDVDSALYGLSIL
NSKQIKLIIHFHEHKGDDYDRLGVNDEYQDILGVKVKURVPIRPNRIAVIEAAA
ANYQIRMSDISEPVDIIIEKRMLESMEKSKI"
2520. 3449
/gene="CPE1005"
CDS
2520. 3449
/gene="CPE1005"
/notes="309 aa, similar to C-terminal of pir:S75981
magnesium/cobalt transport protein sl10507 from
Synecochystis sp. (strain PCC 6803) (387 aa); 22.4%
identity in 263 aa overlap. 2 putative transmembrane
regions were found by PSORT."
/codon_start=1
/transl_table=11
/product="probable divalent cation transporter"
/protein_id="BAB80711.1"
/db_xref="GI:18144666"
/translation="MVIFNIIEENFKVVEEKKDBFPAILMEEDENVLLEKMFSEYCI
NECDYQEQRIDCFKYTFILINIEQVEKQIISKELDIFLGKSPITVYKSNISIL
YDLIEDKEENKCOLFKVDKSPSILYILDRILINNYIIISLETGDKDIELEILKD
PESHLASLLPIREVEVKLRKLIKPIRYIGDTIADIDENEYVIGENKVPKNNINSKIEK
LITSLDLSQELAVRAYSELANKNELMKVFTITAVFLPLELITAFSPMSFESM
PFKTCPGYFYLVTLSMALILLILFKIRKFL"
3516. 4922
/gene="CPE1006"
CDS
3516. 4922
/gene="CPE1006"
/notes="468 aa, similar to pir:E70145 aminopeptidase I
(yscI) homolog (EC 3.4.11.-) from Borrelia burgdorferi
(458 aa); 45% identity in 456 aa overlap"
/codon_start=1
/transl_table=11
/product="probable aminopeptidase I"
/protein_id="BAB80712.1"
/db_xref="GI:18144667"
/translation="MKQDVVLKKYENAWTKYEEADLKQVPSLSDRYRPMFVAKTER
ECVQLANAEKGFKNFYEVINKGEKVTAGDKLYISNMDKTTILIKVGEPLNGLR
IIGAHIDSPMDKVNPLYEDSGIALLDTHYGGKQYOWTPTLAIHGVVKKDQGTIR
VDIKIGDENDPVIGISDLLIHLSADQLDKKAKVVEGEDLINLVGSMPLKGTKEEA
VKANILVLNEKYGITIEDFVSALEVVPAKARDYGLDRSMILAYGHDDRICAYTSA
EALMDLNVKTCVALLVDKBEIGSVGATGQGRFFENIIAEIWDKRGVEYSELKRRC
LQSMMLSAADVTAPDENYPSVCEKNTAYFGHGVFESKVTGARGAGCNDANAEVIA
HLRNIMDKNGVWQTGELGKVDQGGGTIIAYIAQYNNVIDCGVALQNNHAPLEVAS
KADLYETKKCKYAPFEEA"
5047. 5634
/gene="CPE1007"
CDS
5047. 5634
/gene="CPE1007"
/notes="195 aa, partially similar to gp:B90740.3
Hypothetical 24.0K protein from Escherichia coli (476 aa);
27.6% identity in 170 aa overlap. Putative N-terminal
signal sequence and 1 putative transmembrane region were
found by PSORT"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB80713.1"
/db_xref="GI:18144668"
/translation="MNKVLGNLIFLSFILLVTLIFTRYNGENIILLISLIFLRII
FITFEFKYINLGAIKNRSIFSRHKNKNKFNKLVLEKLDPLSGNVFRRFKNKL
BMWKEVTLFLIDIRKFSNLINDALRGKGDEVLRFEFGNGLNNISKSLASKNLYGVGG
EEFLVANSKNKEBIIIEVLNKLNDLKKIRLSLNK"
5707. 6570
/gene="CPE1008"
CDS
5707. 6570
/gene="CPE1008"

```

```

/notes="287 aa, similar to gp:BCE243712.8 YkOW protein from
Bacillus cereus (892 aa); 30.9% identity in 278 aa
overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB80714.1"
/db_xref="GI:18144669"
/translation="MFFAKERAKEDFIYALEYTEELSRKNNLIIRDKDALINK
KLYTYQPISSLDKIRGYETLRWNHETIGSASDIVKLEEDIDAKLDLYVE
STCFQSLRIQNNIELQSVNISITLHRGVGNLYKISKYNNKPNITLLEENA
DLEVRGNAINEIILKKYGFKISIDDFGKYSSINRVLKIPDQIKIKPQFISDIPNS
KYIAJINSISNFAKTLGAEVLIEGVESKEYYDFPKLLDFDVIQGFYFSKALNEKEFLE
YNNLGVSVSI"
6738. 7097
/gene="mgSA"
CDS
6738. 7097
/gene="mgSA"
/notes="119 aa, similar to sp:MGSA_BACSU METHYLGLYOXAL
SYNTHASE (EC 4.2.99.11) (MGS) from Bacillus subtilis (137
aa); 62.9% identity in 116 aa overlap
CPE1009"
/codon_start=1
/transl_table=11
/product="methylglyoxal synthase"
/protein_id="BAB80715.1"
/db_xref="GI:18144670"
/translation="MKIALIADHKKKEIMELAKDFDKLSKHILVATGTTGLKIMQN
TSLEVKRKSPLGQDQIEIGANVHDVDMVIFLDRDPLTAQPHPEPDISALLRLCDVYK
VPLANTBSAKLIMADI"
7278. 7613
/gene="CPE1010"
CDS
7278. 7613
/gene="CPE1010"
/notes="111 aa, no significant homology. Putative
N-terminal signal sequence and 2 putative transmembrane
regions were found by PSORT."
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB80716.1"
/db_xref="GI:18144671"
/translation="NRGNINLQMSNWIIIVSTILTITKYILENLIELCKFNKKIGLTTX
KIRKIVYIFILTVISVIIIGITTFYKRWFLNAVSIQNRILIVAFQTQIFGLIITLF
INRKYYKKV"
7739. 7999
/gene="CPE1011"
CDS
7739. 7999
/gene="CPE1011"
/notes="86 aa, no significant homology. Putative N-terminal
signal sequence and 1 putative transmembrane region were
found by PSORT"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

```

Alignment Scores:

Pred. No.:	5,128-43	Length:	298050
Score:	703.50	Matches:	147
Percent Similarity:	67.53%	Conservative:	61
Best Local Similarity:	47.73%	Mismatches:	97
Query Match:	46.34%	Indels:	3
DB:	1	Gaps:	3

US-09-308-397-2 (1-306) x AP003189 (1-298050)

QY	1	MetThrLysThralapheLeuPheAlaGlyGingIyAlaGlnTyrLeuGlyMetGlyArg 20
Db	68429	ATGGCTAAACTAGGATTTTATTTCAGGCGCAAGAGCCCAATATGTAGGAATGGCAAA 68488
QY	21	AspPheTyrAspGlnTyrProIleValIysGluThrIleAspArgAlaSerGlnValIeu 40
Db	68489	GAATTCCTTTGATAACTTTTGAGAGAAAGTAAAGAGGTATTTTAAAGAAAGTAGTGAAGCCCTTA 68548

```
QY 41 GlyTyrAspLeuArgTyrLeuIleAspThrGlu---GluAspLysLeuAsnGlnThrArg 59
Db 68549 GGCATAGATATGAAGAACCTTTGTTTAGTGATCCTGGAGGATTATTAAATAAACTGAA 68608
QY 60 TyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnLys 79
Db 68609 TTTACTCAGCCTGCAATAATAACCAAACTAGGCTATATCTACAGCACTAGATAAATTA 68668
QY 80 GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAla 99
Db 68669 GGAGTTAAATCTCATATAGTTGCGATTAGCTTAGGAGATAATTCGCACCTATTTCAT 68728
QY 100 SerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMet 119
Db 68729 AGTGGTGCCCAATTAATTTGAAGATGCGATTAACTAGTTAAAAAAGAGGTAAATTCATG 68788
QY 120 GluGluAlaAlaProAlaAspSerGlyLysMetValAlaValAlaLeuAsnThrProValGlu 139
Db 68789 CAAAGAGCTGTGCAGAGGTATGTTGGATGGTTCAGTTTAAAGATGACTCCAGAA 68848
QY 140 ValIleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaSerTyr 159
Db 68849 CAAATTTGATGAAATAATAAGAAAAAGTTCTCCATATGGAATAGTTGAGGGTGCAATAT 68908
QY 160 AsnThrProAlaGlnIleAlaGlyGluValValAlaValAlaValAspArgAlaValGlu 179
Db 68909 AATTCACCTGGCCAAATTTGTTATTTCTGGAGAATTAGTAGCTTTAGAAAAAGCTATGGA 68968
QY 180 LeuLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHis 199
Db 68969 TTTATAAAGAGCTGTGGTGA--AGAGCTATAAAATTACCTGTATCAGCTCCATTTCAC 69025
QY 200 ThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPhe 219
Db 69026 TGTTCAATGTTCAACACGAGCTGAGAAATTAGAGATGAATTAATAAATTCATA 69085
QY 220 SerAspPheThrCysProLeuValGlyAsnThrGluAla---AlaValMetGlnLysGlu 238
Db 69086 AATAAATTAATGGAATTTGTTATGTTCTATGTTAAGGAGAGAGCTATTAGAGATGAT 69145
QY 239 AspIleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIle 258
Db 69146 AATATAATAGAACTTTTAACTTCTCAAGTTAAAGAGCCTGTTTATTATAATGATATA 69205
QY 259 GlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeu 278
Db 69206 GAAAAAATGATTGAAGTGTGTAGATACCTTTTATTGAAATAGTCTCTGTAAGCTTTA 69265
QY 279 SerGlyPheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAla 298
Db 69266 AGTGGATTTGTTAAAAAATAAATAAAATGTTACAGTGTCTTAATGTAGAGGATTITAAA 69325
QY 299 SerLeuValAlaLeuLeuGluLys 306
Db 69326 TCITTAGAAAAAACATTATCAAA 69349
```

Search completed: June 11, 2003, 21:02:50
Job time : 2443 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 11, 2003, 20:05:52 ; Search time 245 Seconds
(without alignments)
2812.700 Million cell updates/sec

Title: US-09-308-397-2

Perfect score: 1518
Sequence: 1 MTKTAFLFAGGAGYLGWGR.....QTAHLAHDQASIVALLEX 306

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlsp
-Q=/cgn2.1/USPTO_epool/US09308397/runat_06062003_112348_28863/app_query.fasta_1.455
-DB=N Geneseq_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pfco -NOEM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09308397@cgn1.1.208@runat_06062003_112348_28863 -NCPUS=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -HEADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq_101002.*
1: /SID82/gcgdata/geneseg/genesegq-embl/NA1980.DAT.*
2: /SID82/gcgdata/geneseg/genesegq-embl/NA1981.DAT.*
3: /SID82/gcgdata/geneseg/genesegq-embl/NA1982.DAT.*
4: /SID82/gcgdata/geneseg/genesegq-embl/NA1983.DAT.*
5: /SID82/gcgdata/geneseg/genesegq-embl/NA1984.DAT.*
6: /SID82/gcgdata/geneseg/genesegq-embl/NA1985.DAT.*
7: /SID82/gcgdata/geneseg/genesegq-embl/NA1986.DAT.*
8: /SID82/gcgdata/geneseg/genesegq-embl/NA1987.DAT.*
9: /SID82/gcgdata/geneseg/genesegq-embl/NA1988.DAT.*
10: /SID82/gcgdata/geneseg/genesegq-embl/NA1989.DAT.*
11: /SID82/gcgdata/geneseg/genesegq-embl/NA1990.DAT.*
12: /SID82/gcgdata/geneseg/genesegq-embl/NA1991.DAT.*
13: /SID82/gcgdata/geneseg/genesegq-embl/NA1992.DAT.*
14: /SID82/gcgdata/geneseg/genesegq-embl/NA1993.DAT.*
15: /SID82/gcgdata/geneseg/genesegq-embl/NA1994.DAT.*
16: /SID82/gcgdata/geneseg/genesegq-embl/NA1995.DAT.*
17: /SID82/gcgdata/geneseg/genesegq-embl/NA1996.DAT.*
18: /SID82/gcgdata/geneseg/genesegq-embl/NA1997.DAT.*
19: /SID82/gcgdata/geneseg/genesegq-embl/NA1998.DAT.*
20: /SID82/gcgdata/geneseg/genesegq-embl/NA1999.DAT.*
21: /SID82/gcgdata/geneseg/genesegq-embl/NA2000.DAT.*
22: /SID82/gcgdata/geneseg/genesegq-embl/NA2001A.DAT.*
23: /SID82/gcgdata/geneseg/genesegq-embl/NA2001B.DAT.*
24: /SID82/gcgdata/geneseg/genesegq-embl/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	1518	100.0	921	19	AAV37239
2	1518	100.0	921	22	AAH90730
3	1518	100.0	19702	19	AAV52140
4	1515	99.8	945	22	AAH90843
5	1515	99.8	3580	19	AAV52423
6	1514	99.7	921	23	AAV53845
7	1048	69.0	2365589	24	ABA90521
8	1036	68.2	942	24	ABN68641
9	996.5	65.6	924	24	ABN68640
10	996.5	65.6	2155561	24	ABN71527
11	937	61.7	942	23	AAV53186
12	920	60.6	3656	20	AAV13514
13	863	56.9	547	19	AAV37241
14	858	56.5	1196	19	AAV98542
15	858	56.5	1199	18	AAV30801
16	858	56.5	1209	19	AAV37406
17	671.5	44.2	636	18	AAV30880
18	643.5	42.4	930	23	AAV53191
19	636	41.9	939	23	AAV53241
20	636	41.9	1830121	17	AAV42063
21	627.5	41.3	950	24	ABK73852
22	627.5	41.3	960	24	ABK73795
23	613	40.4	450	19	AAV37240
24	594	39.1	954	24	ABQ90247
25	589.5	38.8	2301	24	ABQ70706
26	588.5	38.8	2944528	24	ABA03041
27	586	38.6	319630	24	ABQ67194
28	586	38.6	3011208	24	ABQ69245
29	566	37.3	46593	21	AAH81456
30	566	37.3	349980	21	AAF21612
31	566	37.3	837096	21	AAH81489
32	556.5	36.7	906	23	AAV51823
33	556.5	36.7	936	23	AAV54395
34	556.5	36.7	939	23	AAV54159
35	550.5	36.3	1170	21	AAZ51219
36	547.5	36.1	927	19	AAV31440
37	547.5	36.1	999	22	AAV06204
38	528	34.8	273254	21	AAH81914
39	528	34.8	1230025	20	AAV91990
40	515.5	34.0	936	24	ABN93524
41	510	33.6	31940	22	AAF28526
42	479.5	31.6	41599	22	AAI66165
43	474.5	31.3	6251	18	AAV74336
44	454.5	29.9	1038602	20	AAZ01425
45	439.5	29.0	1357	21	AAV49129

ALIGNMENTS

RESULT 1
AAV37239
ID AAV37239 standard; DNA; 921 BP.
XX
AC AAV37239;
XX
XX 24-SEP-1998 (first entry)
DT
DE DNA encoding a FabD polypeptide.
XX
XX FabD protein; malonyl-CoA:ACP family; diagnosis; infection; vaccine;
screen; ss.
XX Streptococcus pneumoniae.
XX
XX Key Location/Qualifiers
FH 1..921
CDS /tag= a

```

FT      /product= FabD
PN      WO3822133-A1.
XX      28-MAY-1998.
PD      14-NOV-1997; 97WO-US20992.
XX      18-NOV-1996; 96US-0031180.
PR      (SMIK) SMITHKLINE BEECHAM CORP.
XX      Gentry DR, Lonsdale JT, Payne DJ, Pearson SC, Van Aller G;
XX      WPI; 1998-312173/27.
DR      P-PSDB; AAM60856.
XX      New isolated Streptococcus pneumoniae FabD gene - used to develop
PT      products for the diagnosis, prevention and treatment of bacterial
PT      diseases, particularly S. pneumoniae infection
XX      Claim 6; Pages 5-6; 45pp; English.
XX      The present sequence encodes a FabD protein of Streptococcus pneumoniae
CC      010093 (NCIMB 40800). The novel FabD polypeptides are related to other
CC      proteins of the malonyl-CoA:ACP family. The products can be used for the
CC      diagnosis of Streptococcus pneumoniae infections. Vectors containing
CC      the FabD DNA sequence can be administered directly to a mammal to
CC      produce the FabD peptide to provoke an antibody/T-cell response in
CC      order to prevent a disease. The peptide can be used to screen for
CC      compounds which modulate its activity.
XX      Sequence 921 BP; 247 A; 193 C; 230 G; 251 T; 0 other;
SQ

Alignment Scores:
Pred. No.: 7 34e-148 Length: 921
Score: 158.00 Matches: 306
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0

US-09-308-397-2 (1-306) X AAV37239 (1-921)

Qy      1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyValGlnTyrLeuGlyMetGlyArg 20
Db      1 ATGACTAAACAGCCCTTTTATTTGCTGTCGTCAGAGTCCCGAGTATAGGATGGACGG 60
Qy      21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
Db      61 GATTTCATGATCAGTATCCGATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 120
Qy      41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60
Db      121 GGTTATGATTACGTTATCTCATCGATACGGAAGAGACAACTCAATCAGACCCGCTAT 180
Qy      61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGlnGlnLysGly 80
Db      181 ACGCAACAGCCATCTAGGATCTGGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 240
Qy      81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
Db      241 TATCAGCCTGATATGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 300
Qy      101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
Db      301 GCGGCTTGATTGTAAGTATGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Qy      121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
Db      361 GRAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Qy      141 IleGluAlaCysGlnLysAlaSerGluLeuGlyValThrProAlaAsnTyrAsn 160

```

CC The present invention relates to nucleic acids (AAH90701-AAH90918) encoding polypeptides (AAH01002-AAH0114), which are essential for the viability of a bacterial cell wall. The acronym CFE stands for "CEG For Expression", where CEG stands for "Conserved Essential Gene". The nucleic acids are useful for detecting the presence of proteins essential for the viability of a bacterial cell wall in samples such as cells, tissues, biological fluids, blood, serum, nose, ear or throat swabs with ligands, and for detecting corresponding target nucleic acid molecules with complementary sequences. The nucleic acids are also useful for determining whether a genomic nucleotide sequence of interest is essential for viability of a bacterial cell or whether it resides within an operon, by integrating an exogenous nucleotide sequence comprising a portion of an open reading frame of the genomic sequence of interest (comprising 200-500 base pairs) into the genomic sequence of interest which confers a selectable phenotype to the cell, and determining cell viability with a selection agent such as chloramphenicol. The nucleic acids and proteins are also useful as vaccines and for treating bacterial infections with gene therapy and antisense therapy. The nucleic acids also enable identification of targets suitable for the treatment of antibiotic resistant bacterial infections.

SQ Sequence 921 BP; 246 A; 193 C; 232 G; 250 T; 0 other;

Alignment Scores:

Pred. NO.: 7.34e-148 Length: 921
 Score: 1518.00 Matches: 306
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0

US-09-308-397-2 (1-306) x AAH90730 (1-921)

QY 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
 DB 1 ATGATCAAAACAGCCCTTTTATTCCTGCTCAAGTGCCAGATCTAGGAGTGGGACGG 60
 QY 21 AspPheTyrAspGlnTyrProileValLysGluThrIleAspArgAlaSerGlnValLeu 40
 DB 61 GATTTCCTATGATCAGTATCCGATTCTTAAAGAAACGATTGATCGAGCGAGTCAGGTGTC 120
 QY 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60
 DB 121 GATTGATTTTCGTTATCTATCATGATACGGAAGAGCAAACTCAATCAGCCCGCTAT 180
 QY 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLysGly 80
 DB 181 ACGCAACAGCCATTCAGCGACTTCGCTTGTATCTACCGTTTATTGCAAGAAAGGCG 240
 QY 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
 DB 241 TATCAGCCGTGATGCTGCTGGTGGTGTCTCTTGGAGAAATCTCTGCTGGTGGCAAGC 300
 QY 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
 DB 301 GCGCGCTTGGATTGAGATCGGCTTGGTGTAGCTAGCGTGGAGCTTATATGGA 360
 QY 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuValLeuAsnThrProValGluVal 140
 DB 361 GAAGCGGCTCTCTGCTGCTGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTC 420
 QY 141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
 DB 421 ATTGAAGAAGCCCTGTCAAAAGCTTCTGAACTTGGAGTGGTGTATCTCCAGCCCACTATAAC 480
 QY 161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAlaAspArgAlaValGluLeu 180
 DB 481 ACACCTGCACAAATCGTCTGCTGAGAGTGGTGTGAGTTGATCGAGCGGTTGAACCT 540
 QY 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr 200
 DB 541 TTGCAAGAAGCAGGTCGCAACGCTTGATCTCTTAAAGGTGTCAGGTCCCTTCACACC 600

QY 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
 DB 601 GCTCTCTTTCAGCCAGCTAGCCAGAACTAGCTGAACCTCTAGCTCAGGTAAGTTTTC 660
 QY 221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle 240
 DB 661 GATTTACTTGTCCCTAGTCGCGCATCAGAACTGCTGTGATGCAAAAGAGGACATT 720
 QY 241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
 DB 721 GCTCAGCTCTTACGCGCTCAGGTCAGGAACCCGTTCTGTTCTATCAAAAGTATTGGGTC 780
 QY 261 MetGlnGluAlaGlyLysSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280
 DB 781 ATGCAGAGAGCAGGCAATAGCACTTATTCGAGATGGACCGGGGAAAGTCTTGTGAGT 840
 QY 281 PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
 DB 841 TTGTGTAATAAATTGATCAAACTGCTCACTTAGCTCATGTGTAAGATCAAGCGAGTTTA 900
 QY 301 ValAlaLeuLeuGluLys 306
 DB 901 GTAGCACTTTTAGAAAAA 918
 RESULT 3
 AAV52140/c
 ID AAV52140 standard; DNA; 19702 BP.
 XX
 AC AAV52140;
 DT
 XX 23-OCT-1998 (first entry)
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:7.
 KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical composition; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO98:8931-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-US19588.
 XX
 PR 31-OCT-1996; 96US-0029960.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XX
 WPI; 1998-272225/24.
 XX
 PT Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PT pneumoniae
 XX
 PS Claim 1; Page 194-205; 1409pp; English.
 XX
 CC The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment or a sequence at least 95%
 CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
 CC library which contain sequences that hybridize to the target sequence and

CC isolating the nucleic acid molecules from the members; or (b) isolating
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
 CC molecules whose nucleotide sequence is homologous to amplification
 CC primers derived from the fragment of the S. pneumoniae genome to prime
 CC the amplification and isolating the amplified sequences. The computer
 CC readable medium can be used in a computer-based system for identifying
 CC fragments of the S. pneumoniae genome of commercial importance, or
 CC expression modulating fragments of the S. pneumoniae genome. Products
 CC from the present invention can be used in diagnosis kits and assays, and
 CC pharmaceutical compositions and vaccines for S. pneumoniae.

XX Sequence 19702 BP; 5736 A; 4548 C; 3728 G; 5690 T; 0 other;

Alignment Scores:

Pred. No.: 4,656-146 Length: 19702
 Score: 1518.00 Matches: 306
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 19 Gaps: 0

US-09-308-397-2 (1-306) x AAV52140 (1-19702)

QY 1 MetThylsThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
 Db 17149 ATGACTAAACAGCCTTTTATTGCTGCTCAGGTGCCAGTATCTAGGATGGACGG 17090
 QY 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
 Db 17089 GATTCTATGATCAGTATCCGATGTTTAAAGAACGATTGATCGACGACCGAGTGTCTC 17030
 QY 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60
 Db 17029 GGTATGATTGGTTATCTCATCGATACGGAAGAAGACAACTCAATCAGACCCGCTAT 16970
 QY 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGlnGlnLysGly 80
 Db 16969 ACCGACCCAGCCTTCTAGGACTTCGGTTGCTATCTACCGTTTATGCAAGAAAGGCG 16910
 QY 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
 Db 16909 TATCAGCCTGATATGTTGCTGTTGTTGCTCTTGGAGATACTCTGCTTGTGGCAAGT 16850
 QY 101 GlyAlaIleAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
 Db 16849 GCGCCTTGATTTGAAGATCGGTTGCTTGGTAGCTAAGCTGGAGCCTATATGGAA 16790
 QY 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
 Db 16789 GAAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 16730
 QY 141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
 Db 16729 ATTGAAGAAGCCTGTCAAAAAGCTTCTGAACTTGGAGTGGTACTCCAGCCAACTATAAC 16670
 QY 161 ThrProAlaGlnIleValIleAlaGlyGluValAlaValAlaAspArgAlaValGluLeu 180
 Db 16669 ACACCTTGCAAAATCTCTATCTGAGAGAGTGGTTCAGTGTGATCGACGGTGAACCTT 16610
 QY 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr 200
 Db 16609 TTGCAAGAAGCAGGTGCCAAACGCTTGAATCTCTTAAGGTGTGAGTCCCTTTTACACC 16550
 QY 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
 Db 16549 GCTCTCTGTAACCTGCTAGCCAGAAACCTAGCTGGAACCTCTAGCTCAGGTAAATTTTCA 16490
 QY 221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle 240
 Db 16489 GATTTTACTTGTCCCTAGTCGCAATACAGAAGCTGCTGTGATGCAAAAGAGGACATT 16430
 QY 241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGlnSerIleGlyVal 260

Db 16429 GCTCAGCTCTTGACGCTCAGGTCAAGGAACCCGTTGTTTCTATGAAAGTATGGGTC 16370
 QY 261 MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280
 Db 16369 ATGCAAGAAGCAGGCATTAAGCAACTTATCGAGATTGCACCGGGAAGTCTTTGTCAGGT 16310
 QY 281 PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
 Db 16309 TTTGTTAAAAAATGATCAAACTGCTCACTTAGTCTATGTGGAAGATCAAGCGAGTTTA 16250
 QY 301 ValAlaLeuLeuGluLys 306
 Db 16249 GTAGCACTTTTAGAAAAA 16232
 RESULT 4
 AAH90843
 ID AAH90843 standard; DNA; 945 BP.
 AC AAH90843;
 DT 02-OCT-2001 (first entry)
 XX 2CFE 34 coding sequence.
 DE Antibacterial; vaccine; gene therapy; bacterial cell wall viability;
 KW CPE; CEG; Conserved Essential Gene; bacterial infection;
 KW antisense therapy; antibiotic resistance; ds.
 XX Streptococcus pneumoniae.
 CS WO200149721-A2.
 FN 12-JUL-2001.
 PD 29-DEC-2000; 2000WO-US35604.
 PP 30-DEC-1999; 99US-0174089.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 PA Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Bruccoleri RE;
 PI Thanassi JA;
 XX WPI; 2001-496721/54.
 DR Nucleic acids encoding conserved essential genes involved in bacterial
 XX replication which are potential targets for the treatment of antibiotic
 PT resistant bacterial infections -
 PT Claim 8; Fig 55; 380pp; English.
 PS The present invention relates to nucleic acids (AAH90701-AAH90918)
 CC encoding polypeptides (AAH90701-AAH90918), which are essential for the
 CC viability of a bacterial cell wall. The acronym CFE stands for "CEG For
 CC Expression", where CEG stands for "Conserved Essential Gene". The nucleic
 CC acids are useful for detecting the presence of proteins essential for the
 CC viability of a bacterial cell wall in samples such as cells, tissues,
 CC biological fluids, blood, serum, nose, ear or throat swabs with ligands,
 CC and for detecting corresponding target nucleic acid molecules with
 CC complementary sequences. The nucleic acids are also useful for
 CC determining whether a genomic nucleotide sequence of interest is
 CC essential for viability of a bacterial cell or whether it resides within
 CC an operon, by integrating an exogenous nucleotide sequence comprising a
 CC portion of an open reading frame of the genomic sequence of interest
 CC (comprising 200-500 base pairs) into the genomic sequence of interest
 CC which confers a selectable phenotype to the cell, and determining cell
 CC viability with a selection agent such as chloramphenicol. The nucleic
 CC acids and proteins are also useful as vaccines and for treating bacterial
 CC infections with gene therapy and antisense therapy. The nucleic acids
 CC also enable identification of targets suitable for the treatment of
 CC antibiotic resistant bacterial infections.
 XX Sequence 945 BP; 254 A; 205 C; 232 G; 254 T; 0 other;

Alignment Scores:
Pred. No.: 1,56e-147 Length: 945
Score: 1515.00 Matches: 305
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.67% Mismatches: 0
Query Match: 99.80% Indels: 0
DB: 22 Gaps: 0

US-09-308-397-2 (1-306) x AAH90843 (1-945)

Qy 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
Db 1 ATGACTAAACAGCGCTTTTATTTGCTGCTCAAGGTGCCAGTATCTAGGATGGGACGG 60
Qy 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
Db 61 GATTTCATGATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCGAGTCAGGTGCTA 120
Qy 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60
Db 121 GGTATGATTTCGCTATCTCATGATACGGAAGACAACTCAATCAGACCGCTAT 180
Qy 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGlnGlnLysGly 80
Db 181 ACGCAACAGCCATTCTAGCCACTTCGCTTCTATCTACCGTTTATTGCAAGAAAGGCG 240
Qy 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
Db 241 TATCAGCCTGATATGGTTGCTGTTGCTCTTGGAGAACTCTGCTGGTGGGCAAGC 300
Qy 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
Db 301 GCGCGCTTGGATTGTAAGATCGGTTGCCCTGCTAGCTAAGCGTGGAGCTATATGGAA 360
Qy 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
Db 361 GAAGCGGCTCTGCTGATCTCGGAAGATGTAGCAGTTCTCAATACGCCAGTAGAGGTC 420
Qy 141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
Db 421 ATTGAAGAAGCCTGTCAAAAAGCTTCTGAACCTTGGAGTGGTTACTCCAGCCAACTATAAC 480
Qy 161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAlaAspArgAlaValGluLeu 180
Db 481 ACACCTGCACAAATCGTATTGCTGGAGAGTGTTCAGTTGATCGAGCGGTTGAACCT 540
Qy 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr 200
Db 541 TTGCAAGAGCAGGTGCCAAACGCTTGATTCCTCTTAAGGTGTCAGTCCCTTTCACACC 600
Qy 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
Db 601 TCTCTCCTTGACCTGTAGCCAGAACTAGCTGAAACTCTGGCTCAGTAAAGTTTTCATCA 660
Qy 221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle 240
Db 661 GATTTCATGTCCTTCTAGTCCGGAATACAAAGCTCTGTGATGCAAAAGAGGACATT 720
Qy 241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
Db 721 GCTCAGCTCTTGACGCGTCAGGTCAAGGAACCCGCTGCTTCTATGAAATATTGGGGTC 780
Qy 261 MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280
Db 781 ATGCAAGAAGCAGGCATTAAGCAACTTTATCAGATTGGACGGGGAAGCTTGTTCAGGT 840
Qy 281 PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
Db 841 TTTGTAAAAAATTGATCAACTGCCTACCTTACTCTCATGTGGAAGATCAAGCGAGTTTA 900
Qy 301 ValAlaLeuLeuGluLys 306

Db 901 GTAGCACATTTTACAAAAA 918

RESULT 5
AAV65243
ID AAV65243 standard; DNA; 3580 BP.
XX AAV65243;
XX AAV65243;
DT 24-DEC-1998 (first entry)
XX
DE S. pneumoniae gene fragment.
XX
KW Streptococcus pneumoniae protein; recombinant; gene expression;
NM DNA chip; virulence; antibody; infection; detection; treatment; ss.
XX
OS Streptococcus pneumoniae.
XX
PN W09826072-A1.
XX
PD 18-JUN-1998.
XX
PF 09-DEC-1997; 97WO-US22578.
XX
PR 13-DEC-1996; 96US-0036281.
XX
PA (ELIL) LILLY & CO ELI.
XX
PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR;
PI Mills BJ, Norris FH, Peery RB, Rostock PK, Rostock PR;
PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ;
PI Young Bellido ML;
XX
DR WPI; 1998-348529/30.
XX
PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
PT for evaluating gene expression, and identification of virulence
PT genes
XX
PS Claim 1; Pages 96-98; 333pp; English.
XX
CC This sequence is a gene fragment from the Streptococcus pneumoniae
CC genome. The invention provides DNA sequences (AAV65201 to AAV65304) from
CC the Streptococcus pneumoniae genome and corresponding protein sequences
CC (AAW80605 to AAW80728). A recombinant host containing a vector comprising
CC any of the above nucleic acids can be used for the recombinant expression
CC of the protein sequences. The invention also provides a DNA chip having
CC arrayed on it at least 15 base pair fragment of any one or more of these
CC DNA sequences. The DNA chip can be used for evaluating gene
CC expression in S. pneumoniae and for identifying virulence genes in
CC S. pneumoniae. Antibodies that selectively bind to the above proteins or
CC peptide fragments can be used to treat S. pneumoniae infection. The
CC antibodies can also be used to detect S. pneumoniae cells.
XX
SQ Sequence 3580 BP; 1035 A; 666 C; 894 G; 985 T; 0 other;

Alignment Scores:
Pred. No.: 9,45e-147 Length: 3580
Score: 1515.00 Matches: 305
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.67% Mismatches: 0
Query Match: 99.80% Indels: 0
DB: 19 Gaps: 0

US-09-308-397-2 (1-306) x AAV65243 (1-3580)

Qy 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
Db 718 ATGACTAAACAGCGCTTTTATTTGCTGCTCAAGGTGCCAGTATCTAGGATGGGACGG 777
Qy 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
Db 778 GATTTCATGATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCGAGTTCAGGTGCTA 837

Qy 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluAspLysLeuAsnGlnThrArgTyr 60
 Db 838 GGTATGATTGGTGTATCTCATCGATACGGAAGAAGACAACTCAATCAGACCCCGCTAT 897
 Qy 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLysGly 80
 Db 898 ACGCAACCGCCATTCTAGGACCTTCGGTTGGCTATCTACCGTTTATTGCAAGAAAGGCG 957
 Qy 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
 Db 958 TATCAGCCTCATATGGTGTGTGTCTCTCTGGGAATACTCTCGCTTGGTGGCAAGC 1017
 Qy 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
 Db 1018 GGGCCCTTGATTTGAGATGGGTTGCTTGGTAGCTAAGCGTGAGCCCTATATGGA 1077
 Qy 121 GluAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
 Db 1078 GAAGCGGCTCCTGCTGATCTGCAAGATGTTGAGAGTTCTCAATACGCCAGTAGAGTTC 1137
 Qy 141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
 Db 1138 ATTGAAGAGCGCTGCATAAGCTTCGAACCTTGGAGTGGTTACTCCAGCCAACTATAAC 1197
 Qy 161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeu 180
 Db 1198 ACACCTGCACAAATCGTCATTGCTGGAGAAAGTGGTTCAGTTGATCGAGCGGTGAACCT 1257
 Qy 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr 200
 Db 1258 TTGCAAGAGCAGGTGCAACAGCTTGATTCCTCTTAAGGTGTCAAGTCCCTTTCACACC 1317
 Qy 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
 Db 1318 TCCTCTCTGAACCTGCTAGCCGAACTAGCTGAACTCTGGCTCAGGTAGTCTTTTCA 1377
 Qy 221 AspPheThrCysProLeuValGlyAsnThrGluAlaValMetGlnLysGluAspIle 240
 Db 1378 GATTTTACTTGTCCCTTAGTCGCAATACAGAGCTGCTGTGATGCAAAAGAGGACATT 1437
 Qy 241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
 Db 1438 GCTCAGCTCTTGACCGCTCAGGTCAAGAACCCGTTCTTCTATGAAGTATTGGGTC 1497
 Qy 261 MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280
 Db 1498 ATGCAAGAGCAGGCATAGCAACTTTATCGAGATTGGACCGGGAAGTCTTGTCTCAGGT 1557
 Qy 281 PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
 Db 1558 TTTGTTAAAAAATTTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGATTTA 1617
 Qy 301 ValAlaLeuLeuGluLys 306
 Db 1618 GTAGCACTTTTAGAAAA 1635

RESULT 6

AAS55845

ID AAS55845 standard; DNA; 921 BP.

XX AC AAS55845;

XX DT 13-FEB-2002 (first entry)

XX DE Streptococcus pneumoniae DNA for cellular proliferation protein #416.

XX KW Antisense; ds; prokaryotic cellular proliferation gene;

XX OS Streptococcus pneumoniae.

XX PN W0200170955-A2.

XX

PD 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207272P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR P-PSDB; AAU37986.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX Claim 27; Seq ID No 9482; Slipp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence encodes an

CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 921 BP; 246 A; 191 C; 232 G; 252 T; 0 other;

SQ

Alignment Scores:

Pred. No.: 1,91e-147 Length: 921
 Score: 1514.00 Matches: 305
 Percent Similarity: 100.00% Conservatives: 1
 Best Local Similarity: 99.67% Mismatches: 0
 Query Match: 99.74% Indels: 0
 DB: 23 Gaps: 0

US-09-308-397-2 (1-306) x AAS55845 (1-921)

Qy 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20

Db 1 ATGACTAAACAGCCCTTTTATTGCTGGTCAAGGTGCCAGTATCTAGGGATGGGACGG 60

Qy 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40

Db 61 GATTTCATGATCATCGATTGTCAGAAAGACGATTGATCGAGCGAGTCAGGTGCTA 120

Qy 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluAspLysLeuAsnGlnThrArgTyr 60

Db 121 GGTATGATTTCGCTTATCTCATCGATACGGAAGAGCAAACTCAATCAGACCCCGCTAT 180

Qy 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnLysGly 80

Db 191 AGCAACAGCCATCTAGGACCTTCGGTTGCTATCTACCGTTTATTCGAAAGAGGCG 240

QY 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
 DB 241 TATCAGCTGATATGTTGCTGTTTCTTCTGGAGTAATCTCTGCTTGGTGGCAGC 300
 QY 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
 DB 301 GCGCGCTTGGATTTTGAAGATGCGGTTGCCCTTGGTAGCTAAGCGTGGAGCCTATATGGAA 360
 QY 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
 DB 361 GAAGCGGCTCTGCTGATCTGGCAAGATGGTAGCAGTCTCAATAGCCAGTAGAGGTC 420
 QY 141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
 DB 421 ATTGAAGAAGCCTGTCAAAAGCTTCTCAACTTGGAGTGGTTACTCCAGCCCACTATAAC 480
 QY 161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAlaAspArgAlaValGluLeu 180
 DB 481 ACACCTGCACAAATCGTCTGCTGGAGAGTGGTTGCAGTGTGATCGAGCGGTTGAACCTT 540
 QY 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr 200
 DB 541 TTGCAAGAGCAGGTGCCAAACGCTTGATTTCTCTTAAGGTGTGAGGTCCCTTTCACAC 600
 QY 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
 DB 601 GCTCTCTTGAACCTGCTAGCCAGAACTAGCTGAAACTCTAGCTCAGGTAAAGTTTTCAT 660
 QY 221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle 240
 DB 661 GATTTTACTTGTCCCTAGTGGCNAATACAGAGCTGCTGTGATGTCGCAAAAGAGACATT 720
 QY 241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
 DB 721 GCTCAGCTCTTCAACGCTCAGGTCAAGAACCCGTTGCTTCTTCTATGAAAGTATTGGGGTC 780
 QY 261 MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280
 DB 781 ATGCAAGAAGCAGGATTAAGCAACTTTATCGAGATTGGACCGGGGAAAGTTTGTCCAGGT 840
 QY 281 PheValLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
 DB 841 TTTGTTAAAAAATTGATCGAACTCTCCTACCTAGCTCATGTGGAAGATCAAGCAGGTTTA 900
 QY 301 ValAlaLeuLeuGluLys 306
 DB 901 GTAGCACTTTTAGAAAAA 918
 RESULT 7
 ID ABA90521 standard; DNA; 2365589 BP.
 AC ABA90521;
 XX ABA90521;
 AC ABA90521;
 XX ABA90521;
 DT 16-MAY-2002 (first entry)
 DE Genomic sequence of Lactococcus lactis IL1403.
 XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.
 KW Lactococcus lactis IL1403.
 OS Lactococcus lactis IL1403.
 XX Lactococcus lactis IL1403.
 XX FR2807446-A1.
 XX 12-OCT-2001.
 XX 11-APR-2000; 2000FR-0004630.
 XX 11-APR-2000; 2000FR-0004630.
 XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
 PA Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 PI

XX WPI; 2002-043418/06.
 XX New nucleotide sequence useful in the identification of Lactococcus
 PT lacticis and related species -
 XX Claim 1; SEQ ID 1; 2504pp; French.
 XX The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (AB553300-AB555621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO200177334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 1,53e-94 Length: 2365589
 Score: 1048.00 Matches: 208
 Percent Similarity: 80.52% Conservative: 40
 Best Local Similarity: 67.53% Mismatches: 58
 Query Match: 69.04% Indels: 2
 DB: 24 Gaps: 1
 US-09-308-397-2 (1-306) x ABA90521 (1-2365589)
 QY 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
 DB 782872 ATGACTAAACACAGCAATTTTATTTCTCAGGTCAAGGGGCACAAAGCTTGGATGGCAGCT 782931
 QY 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
 DB 782932 GACTTATATGACCAATATCAACAGTAAAGCAACTTTTGTAGTCAAGCAAGTCAAGCTTTA 782991
 QY 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60
 DB 782992 GGNATGATTTGCGACCTTTGATTGATGATGATGATGATGATGATGATGATGATGATGAT 783051
 QY 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGlyGly 80
 DB 783052 ACTCAACCTGCAATTTTAAACACCTCTGTGCTTATTTTACCTTTGTAAAGTGGATGG 783111
 QY 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
 DB 783112 ATTAAACCTGACCTTGTGCTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 783171
 QY 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
 DB 783172 GGAATCATTTGATTTTCAAGAAGCAGTAAAGCTTTAGCTTAAACGCTGGTCAATATATGACA 783231
 QY 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
 DB 783232 GAAGTGCACCGACTGCTTGTGCTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 783291
 QY 141 IleGluGluAlaCysGlnLysAlaSerGluLeu-----GlyValValThrProAlaAsn 158
 DB 783292 ATTGAAGAATTTGCCAGAAAGCGCGCAATTTAAAGCGGCTATTGTGATCCAGCAAT 783351
 QY 159 TyrAsnThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaVal 178
 DB 783352 TATAACACCCCGCACAAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 783411
 QY 179 GluLeuLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPhe 198
 DB 783412 GAGTTGCTAAAGAGCCGAGTTCGTAACCTATTGAATTAAGTTTCAAGACCTTTC 783471
 QY 199 HisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSer 218

Db 783472 CATAACCAATTTTAAACACGATCTGAAAAATTCGCTTTGGAGTTCGATAAAATGAT 783531
 Qy 219 PheSerAspPheThrCysProLeuValGlyAsnThrGluAlaValMetGlnLysGlu 238
 Db 783532 TTTAGACCTTTTGAATACCAATTAACTCAATACAGTGCTAAAGTAATGGAATGAT 783591
 Qy 239 AspIleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIle 258
 Db 783592 GAAGTCAAGGAGCTTTTACGCGCTCAAGTCAATGGAACCACTTCGTTTTTATGAATCGGT 783651
 Qy 259 GlyValMetGlnGluAlaGlyIleSerAsnPheIleGlyProGlyLysValLeu 278
 Db 783652 GAACAATGCAAAACTAGGGGCGACTCGCTTTTATGAAGTTGCTCTGGAGAGTACT 783711
 Qy 279 SerGlyPheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAla 298
 Db 783712 TCAGGTTTCATTAATAAAATGATAAAATGCAGAAATGCTAATGTTGAAAATTTAGCT 783771
 Qy 299 SerLeuValAlaLeuLeuGluLys 306
 Db 783772 TCATTTGAAGCTTTTGATTAAATCAG 783795
 RESULT 8
 ID ABN68641
 AC ABN68641 standard; DNA; 942 BP.
 AC ABN68641;
 XX
 DT 01-JUL-2002 (first entry)
 XX Streptococcus polynucleotide SEQ ID NO 5195.
 DE
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX
 OS Streptococcus pyogenes.
 XX
 FN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 EF 29-OCT-2001; 2001WO-GB04789.
 XX
 FR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tettelin H;
 XX
 XX WPI; 2002-352536/38.
 DR P-PSDB; ABP28010.
 DR
 XX New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 XX Claim 7; Page 3863; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity (I), nucleic acids encoding (I), ABN6044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 SQ Sequence 942 BP; 285 A; 159 C; 217 G; 281 T; 0 other;

Alignment Scores:
 Pred. No.: 6.63e-98 Length: 942
 Score: 1035.00 Matches: 209
 Percent Similarity: 81.05% Conservative: 39
 Best Local Similarity: 68.30% Mismatches: 58
 Query Match: 68.25% Indels: 0
 DB: 24 Gaps: 0

US-09-308-397-2 (1-306) x ABN68641 (1-942)

Qy 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
 Db 7 ATGACAAAGACAGACCTTTTATTTGCGCGTCAAGGTCTCAAAATTAGGATGGCAAGG 66
 Qy 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
 Db 67 GATTTTATGATAACTTTGCTATTGTAAGAAAACCTTTGATCAAGCTAGTCAAGTATTG 126
 Qy 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluLysLysLeuAsnGlnThrArgTyr 60
 Db 127 GGATACGATTTGCGCGCTTTGATTGATAGTACAGATTAAACTTAATCAAACTAGCTAT 186
 Qy 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnLysGly 80
 Db 187 ACGAACACAGCTATTTGACATCAATCAATGCTATTACCGTGTGTTGGCTTACATCAC 246
 Qy 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
 Db 247 GTTAAACCGGATATGGTAGCTGTCTTCTTAGAGATATCTCAGCTTTGGTAGCATCA 306
 Qy 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
 Db 307 GGGGCACTCTCTTTGAAGATACCTTATCTTAGTAGTAGAGAGCCGCTTGATGGAG 366
 Qy 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
 Db 367 GAAGCAGCACCAAGGATCTGGAAGATGTTTCCCGTTATGAATACAGACGTCGAAGTC 426
 Qy 141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
 Db 427 ATCGAAGAAGTCTGCTCAAAATAGCTGCTAAACATCGAGTGGTGTCTCCAGCAAACTAAT 486
 Qy 161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeu 180
 Db 487 ACTCTAGTCAATTTGAATTTGGTGGTCAGACAGATGCTGTGAACGGCGAGTTGAACCT 546
 Qy 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr 200
 Db 547 TTAAGGAAGAGGGAGTTAAGCGTTTAACTCCCTTAAACGCTGTCAGTCTCTTCCACACT 606
 Qy 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGlnThrLeuAlaGlnValSerPheSer 220
 Db 607 CTTTGTGTAGAACCAAGCTAGCCGCTTGTGGCTAAAGAGTTGGAAGATACCAACTTCAGT 666
 Qy 221 AspPheThrCysProLeuValGlyAsnThrGluAlaValMetGlnLysGluAspIle 240
 Db 667 GACTTCAAGATTCCTTTGGTTGGTAATACCGAAGCTAATATTATGAAAAAGACCGGTATC 726
 Qy 241 AlaGlnLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
 Db 727 CCAGAACTATTAGCCCGCTCAAGTCAATGAGCCCTTTCGTTTATGACAGTGTTCGCACT 786

Qy	261	MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly	280
Db	787	TTTGTAGAAAGTGGCATACACAAATTTCATTGAGGTAGGACCAGGTAAAGTTTTTCACAGGT	846
Qy	281	PheValIlyLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu	300
Db	847	TTTGTGAGAAATTTGATAAAATTTACTATGTTACTAGTGTGTGAAAACATGGTAAAGCTTA	906
Qy	301	ValAlaLeuLeuGluLys	306
Db	907	CGTTATTTTATAGATAGA	924
RESULT 9			
ABN68640			
ID	ABN68640	standard; DNA; 924 BP.	
XX	AC	ABN68640;	
XX	DT	01-JUL-2002 (first entry)	
XX	XX	Streptococcus polynucleotide SEQ ID NO 5193.	
XX	DE		
XX	XX	Streptococcus; GAS; CBS; group B streptococcus; Streptococcus agalactiae;	
KW	KW	group A streptococcus; Streptococcus pyogenes; antibacterial; gene;	
KW	KW	antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.	
XX	XX		
OS	OS	Streptococcus agalactiae.	
XX	XX		
PN	PN	WO200234771-A2.	
XX	XX		
PD	PD	02-MAY-2002.	
XX	XX		
PF	PF	29-OCT-2001; 2001WO-GB04789.	
XX	XX		
PR	PR	27-OCT-2000; 2000GB-0026333.	
PR	PR	24-NOV-2000; 2000GB-0028727.	
PR	PR	07-MAR-2001; 2001GB-0005640.	
XX	XX		
PA	PA	(CHIR-) CHIRON SPA.	
PA	PA	(GENO-) INST GENOMIC RES.	
XX	XX		
PI	PI	Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;	
PI	PI	Tetzelin H;	
XX	XX		
DR	DR	WPI; 2002-352536/38.	
DR	DR	P-PDB; ABP28005.	
XX	XX		
PS	PS	New Streptococcus protein for the treatment or prevention of infection	
PT	PT	or disease caused by Streptococcus bacteria, such as meningitis, and	
PT	PT	for detecting a compound that binds to the protein -	
XX	XX		
PS	PS	Claim 7; Page 3863; 4525pp; English.	
XX	XX		
CC	CC	The invention relates to a protein (ABP5413-ABP30895) from group B	
CC	CC	(Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS	
CC	CC	(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in	
CC	CC	the specification. The proteins have antibacterial and antiinflammatory	
CC	CC	activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and	
CC	CC	antibodies that bind (I) are used in the manufacture of medicaments for	
CC	CC	the treatment or prevention of infection or disease caused by	
CC	CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.	
CC	CC	Nucleic acids encoding (I) are used to detect Streptococcus in a	
CC	CC	biological sample. (I) is used to determine whether a compound binds to	
CC	CC	(I). A composition comprising (I) or a nucleic acid encoding (I), may be	
CC	CC	used as a vaccine or diagnostic composition. The disease caused by	
CC	CC	Streptococcus that is prevented or treated may be meningitis. Nucleic	
CC	CC	acid encoding (I) may be used to recombinantly produce (I) and may be	
CC	CC	used in gene therapy. Antibodies to (I) are used for affinity	
CC	CC	chromatography, immunoassays, and distinguishing/identifying	
CC	CC	Streptococcus proteins.	
XX	XX		
SQ	Sequence	924 BP; 311 A; 132 C; 194 G; 287 T; 0 other;	

Alignment Scores:
 Pred. No.: 7,99e-94 Length: 924
 Score: 996.50 Matches: 203
 Percent Similarity: 79.08% Conservative: 39
 Best Local Similarity: 66.34% Mismatches: 63
 Query Match: 65.65% Indels: 1
 DB: 24 Gaps: 1

US-09-308-397-2 (1-306) x ABN68640 (1-924)

Qy	1	MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg	20
Db	1	ATGAATAAAGTTTTCATTTCTATTTCCTGGTCAAGTGCTCAAAAGTTAGGAATGGCTAGA	60
Qy	21	AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu	40
Db	61	GAITTTATATGAACACTTCCCGATTGTTAAAGAAACTTTTGATAGCGCAAGTCATGTTTTA	120
Qy	41	GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr	60
Db	121	GGATATGATTGAGAGAAATTCATTGATAAAGACTTAGACAAATTAATCAGAGAAATAT	180
Qy	61	ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeu---	79
Db	181	AGCGAACAGCTATTCTCAACGACTTCGACTCTCTATTATCGTTTAATCTTAAAGAAAT	240
Qy	80	GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAla	99
Db	241	GAACCTTAGCCAGATATGTTGGCAGACTTCTTTGGGAGATACAGTGCCTTTAGTTGCT	300
Qy	100	SerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMet	119
Db	301	TCAGGTGCTATAAGATTTTGAAGATGCTGTAGTACTTCTGTAGACGAGGCCAACATG	360
Qy	120	GluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGlu	139
Db	361	GAAGCAGCAGCACCTCGCTGGTAGTGGCAAAATGGTAGCAGTTTAAATGCTGATCTCTCA	420
Qy	140	ValIleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyr	159
Db	421	ATAATTGAAGATGCTTGTAAAGAGCCTCTCAGTTTGGTATTGTAGTCTCGCAATAT	480
Qy	160	AsnThrProAlaGlnIleValIleAlaGlyValValAlaValAspArgAlaValGlu	179
Db	481	AATACTCCAAAGCAAAATTTGTATATAGGGGAGAGTCAATCGCTGTAATGCAGCTGTAGAA	540
Qy	180	LeuLeuGlnGluAlaGlyAlaIlyArgLeuIleProLeuLysValSerGlyProPheHis	199
Db	541	GAATTAAACACACAGAGAGTAAACGTTTAACTTCTCTGAATGTTTCAGTCTCTTTTAC	600
Qy	200	ThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPhe	219
Db	601	ACTGCTTTATTGAAACCTGCAAGTCAGAAATTTAGCGATGTTTTAGATAAGTTCATTTT	660
Qy	220	SerAspPheThrCysProLeuValGlyAsnThrGluAlaIleValMetGlnLysGluAsp	239
Db	661	TCAGTTTCTGAAATCCCTGTAAATTTGGAAAAACAGAGCCGAGATATGAAAAAAGCAGAT	720
Qy	240	IleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGly	259
Db	721	ATCAAAATCTTTATTGGCAGCCCAAGTTATGGAACCTGTACGCTTTGATGAGTCTATTGAA	780
Qy	260	ValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSer	279
Db	781	ACCATGAAAAAATCGGAATGACACAGCTGTTGTGAAATTTGGTCCAGGGAGGACTCTCTCA	840
Qy	280	GlyPheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSer	299
Db	841	GGTTCTCTTGAAAAAATAGATAGCTCATTCATCAGTACATCTCTGTGAAGATAAAATGGA	900
Qy	300	LeuValAlaLeuLeuGlu	305
Db	901	TTTAAATAATCTTTAAGGAA	918

antibiotic; antibacterial; drug design.

Enterococcus faecalis.

WO200170955-A2.

27-SEP-2001.

21-MAR-2001; 2001WO-US09180.

21-MAR-2000; 2000US-191078P.

23-MAY-2000; 2000US-206848P.

26-MAY-2000; 2000US-207727P.

23-OCT-2000; 2000US-242578P.

27-NOV-2000; 2000US-253625P.

22-DEC-2000; 2000US-257931P.

16-FEB-2001; 2001US-269308P.

(ELIT-) ELITRA PHARM INC.

Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

Yamamoto RT, Xu HH;

WPI; 2001-611495/70.

P-P8DB; AAU35327.

New polynucleotides for the identification and development of

antibiotics, comprise sequences of antisense nucleic acids.

Claim 27; Seq ID No 6823; 51pp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence encodes an essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part

of the printed specification, but was obtained in electronic

format directly from WIPO at

ftp.wipo.int/pub/published_pct_sequences.

Sequence 942 BP; 300 A; 166 C; 234 G; 242 T; 0 other;

Alignment Scores:

Pred. No.: 1,2e-87 Length: 942
Score: 937.00 Matches: 187
Percent Similarity: 75.17% Conservative: 40
Best Local Similarity: 61.92% Mismatches: 75
Query Match: 61.73% Indels: 0
DB: 23 Gaps: 0

US-09-308-397-2 (1-306) x AAS53186 (1-942)

3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22

16 AAACAGCGATTATTATTTAGTGACACAGGAGCCAGATATCAAGGGATGGGTAAGAATTA 75

23 TyrAspGlnTyrProIleValIysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42

76 TATCACCAGAGCGATTGTCGGGAACCTTCGATGAGCAAGTATCATCTTAGGTAT 135

43 AspLeuArgTyrLeuIleAspThrGluAspLysLeuAsnGlnThrArgTyrThrGln 62

.....

Db 136 GAGATGGCAGAACTTTGTTTACTGAAATGAACGCTTAAATGAACACGAATATACGCAA 195

Qy 63 ProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluIysGlyTyrGln 82

Db 196 CCTGCTATTATTAACAGTCAGTGTGCGCATTTTACCGTCTTTTGCACAAAAGGACTAACG 255

Qy 83 ProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSerGlyAla 102

Db 256 CTGTAGTGTAGCGGGTTTAAGCTTTAGGGGAATATAGCGCTTTGGTTCCAGCGGGGT 315

Qy 103 LeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluGluAla 122

Db 316 TTGCGCTTTTTCAGAACGAGTGGCTTGTCCAAAAGCGGGTCAGTACATGACAGAAGCA 375

Qy 123 AlaProAlaAspSerGlyIysMetValAlaValLeuAsnThrProValGlnValIleGlu 142

Db 376 GCACCAACAAGGAAGTGGCAAAATGGTGTCTGCTCATGAATGCTGAGCGTGAAGTAATGAG 435

Qy 143 GluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThrPro 162

Db 436 AAAGCTGTCCAGAACGAGCTGCTTTCGGAATTTGGCTCCAGCAAAATTAATATACCA 495

Qy 163 AlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeuGln 182

Db 496 CAACAATCTGATGCTGCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 555

Qy 183 GluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeu 202

Db 556 GAAGCTGGTGTGAAGCGAATGATTCGTTAATGTAGTGGCCCTTCCATACGCGCTG 515

Qy 203 LeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPhe 222

Db 616 TTACACACGAGCATCAAAAATTTGGCTCAGGATTTAGCAAAATTTGAACCTTTCAAACGATG 675

Qy 223 ThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIleAlaGln 242

Db 676 CAATTCCTGTCTAGTAAATAGCTGCGCAAAATTTCCCAAGAGCAATTTCAAGCG 735

Qy 243 LeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGln 262

Db 736 TTATTGAAAAGCAAGTCATGCTGCGGTACGTTTGAAGACAGTATCGAAACGATGAAG 795

Qy 263 GluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGlyPheVal 282

Db 796 GCTATGAACGCTGGGAACGATGATGAAGTTGTGTCAGGGGAAACATTAATCGTTTGT 855

Qy 283 LysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeuValAla 302

Db 856 AAAAAAATTGACAAACAACTTGAATGCACCGCTGTGGAGATGTTTGCAACATTACAGAA 915

Qy 303 LeuLeu 304

Db 916 ACOTTA 921

RESULT 12

AA13514

ID AAX13514 standard; DNA; 3656 BP.

XX AAX13514;

AC AAX13514;

DT 19-MAR-1999 (first entry)

XX Enterococcus faecalis genome contig SEQ ID NO:577.

XX Enterococcus faecalis; contig; detection; Enterococcal infection;

KW vaccine; attenuation; computer readable medium; ds.

XX Enterococcus faecalis.

CS WO9850555-A2.

PN 12-NOV-1998.

PD

XX

PF 04-MAY-1998; 98WO-US08985.
 PR 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Barash SC, Dillon PJ, Kunsch CA;
 FI WPI; 1999-045171/04.
 DR New isolated Enterococcus faecalis polynucleotides and polypeptides
 XX - used to develop products for the detection of Enterococcus and for
 PT use in vaccines for prevention or attenuation of Enterococcus
 PT infection.
 XX
 PS Claim 1; Page 1834-1836; 2084pp; English.
 XX
 CC A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC AAX12938 to AAX13919 represent these nucleotide sequences which are
 CC primary nucleotide sequences, also known as contigs. The computer-based
 CC system can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence
 CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.
 XX
 SQ Sequence 3656 BP; 1210 A; 627 C; 837 G; 979 T; 3 other;
 Alignment Scores:
 Pred. No.: 4.35e-85 Length: 3656
 Score: 920.00 Matches: 186
 Percent Similarity: 75.25% Conservative: 39
 Best Local Similarity: 62.21% Mismatches: 73
 Query Match: 60.61% Indels: 1
 DB: 20 Gaps: 0
 US-09-308-397-2 (1-306) x AAX13514 (1-3656)
 QY 3 LysThrAlaPheLeuPheAlaGlyGlnGlyValaGlnTyrLeuGlyMetGlyArgAspPhe 22
 DB 2753 AARACGCGATTTTATTAGTGACAGAGAGCCAGTATCAAGGGATGGGTGAAGAATTA 2812
 QY 23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
 DB 2813 TATCACCAGAAGCGGATTTGTCGGAACTTTTCGATGAAGCAAGTCATATCTTAGGTTAT 2872
 QY 43 AspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyrThrGln 62
 DB 2873 GAGATGCGACAGACTTTGTTTACTGAATGAACGTTTAAATGAACAGATATACGCA 2932
 QY 63 ProIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGlyTyrGln 82
 DB 2933 CCTGCTATTTTAAACAGTCAGTGTGCGCATTTTACCGTCTTTTGGCAAAAAGACTAACG 2992
 QY 83 ProAspMetValAlaGlyLeuSerLeuGlyGlnTyrSerAlaLeuValAlaSerGlyVala 102
 DB 2993 CCTGATGTCGTAGGGGTTTAAAGTTAGGGGATACAGTCTTTGGTTCACCGGGGT 3052
 QY 103 LeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluGluAla 122
 DB 3053 TTGGCGCTTTTCAAGACAGTGGCTTGGTCCAAAAGCGGTCAGTACATGACAGAAGCA 3112
 QY 123 AlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGlnValIleGlu 142
 DB 3113 GCACACACAGGAAGCAAAATGTTGCTGTATGAATGCTGAGCGTGAAGTAATGAG 3172

QY 143 GluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThrPro 162
 DB 3173 AAGCGCTGCCAAGAGCCAGTCTTTCGGAATTCGCTCCAGCAATTAATATACACCA 3232
 QY 163 AlaglnIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeuGln 182
 DB 3233 CAACAATCGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3292
 QY 183 GluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeu 202
 DB 3293 GAAGCTGGTGAAGCGAATGATTCGGTAAATGTGAGTGGCCCTTTCCATACGGCGCTG 3352
 QY 203 LeuGluProAlaSerGlnLysLeuAlaGlnThrLeuAlaGlnValSerPheSerAspPhe 222
 DB 3353 TTACACCAAGCATCAAAAAAATTCGCTCAGGATTTAGCAAAATGAAATTTTCAACACGATG 3412
 QY 223 ThrCysProLeuValGlyAsnThrGluAlaValMetGlnLysGluAspIleAlaGln 242
 DB 3413 CAATTCCTGTCTATGATATACGACTGCCGAATTTATGCCCAAGAGGCAATTCACGCG 3472
 QY 243 LeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGln 262
 DB 3473 TTATTGAAAAAGCAAGTCATGTCGCGTACGTTTTCGAAGACAGTATCGAAACGATGAAG 3532
 QY 263 GluAlaGlyIleSerAsnPheIleGluIle-GlyProGlyLysValLeuSerGlyPheVa 282
 DB 3533 GCTATGAACGTAAGAACGATGTTGAAGTTNGTCCAGGGAACATTAACYGTTTGT 3592
 QY 282 llyLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
 DB 3593 TAAAAAATTCACCAACCAATTAATGAATGACCGTGTGAGATGTTGCCACATTA 3647

RESULT 13
 AAV37241
 ID AAV37241 standard; DNA; 547 BP.
 XX
 AC AAV37241;
 XX
 DT 24-SEP-1998 (first entry)
 XX
 DE C-terminal coding region encoding a FabD polypeptide embodiment.
 XX
 KW FabD protein; malonyl-CoA:ACP family; diagnosis; infection; vaccine;
 screen; ss.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO9822133-A1.
 XX
 PD 28-MAY-1998.
 XX
 PF 14-NOV-1997; 97WO-US20992.
 XX
 PR 18-NOV-1996; 96US-0031160.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 PI Gentry DR, Lonsdale JT, Payne DJ, Pearson SC, Van Aller G;
 DR
 DR WPI; 1998-312173/27.
 DR P-PSDB; AAW60858.
 XX
 XX New isolated Streptococcus pneumoniae FabD gene - used to develop
 PT products for the diagnosis, prevention and treatment of bacterial
 PT diseases, particularly S. pneumoniae infection
 XX
 PS Disclosure; Pages 6-7; 45pp; English.
 XX
 CC The present sequence represents the C-terminal coding region of FabD
 CC embodiment (coding begins at nucleotide 2) of Streptococcus pneumoniae
 CC 010993 (NCIMB 40800). The novel FabD polypeptides are related to other
 CC proteins of the malonyl-CoA:ACP family. The products can be used for the

CC diagnosis of Streptococcus pneumoniae infections. Vectors containing
 CC the FabD DNA sequence can be administered directly to a mammal to
 CC produce the FabD peptide to provoke an antibody/T-cell response in
 CC order to prevent a disease. The peptide can be used to screen for
 CC compounds which modulate its activity.

XX SQ Sequence 547 BP; 154 A; 115 C; 132 G; 146 T; 0 other;

Alignment Scores:

Pred. No.: 2,678-80 Length: 547
 Score: 863.00 Matches: 176
 Percent Similarity: 97.25% Conservative: 1
 Best Local Similarity: 96.70% Mismatches: 5
 Query Match: 56.85% Indels: 0
 DB: 19 Gaps: 0

US-09-308-397-2 (1-306) x AAV37241 (1-547)

QY 125 AlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGluGluAla 144
 Db 2 GCTGAACCTGGCAAGATGGTAGCAGTCTCTAATACGCCAGTAGAGGTCATTGAAGAAGCC 61
 QY 145 CysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnThrAsnThrProAlaGln 164
 Db 62 TGTCAAAAGCTCTGGAACTTGGAGTGGTTACTCCAGCCAACTATAACACACCTGCACAA 121
 QY 165 IleValIleAlaGlyGluValValAlaValAlaAspArgAlaValGluLeuGlnGluAla 184
 Db 122 ATCGTCATTGCTGGAGAAGTGGTTGCAGTTCATCGAGCGGTTGAACCTTTGCAGAAGCA 181
 QY 185 GlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeuLeuGlu 204
 Db 182 GGTGCCAAACCTTGATTCCTTTAAGGTGTCAAGTCCCTTTACACCGCTCTCTTGAG 241
 QY 205 ProLaserGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPheThrCys 224
 Db 242 CCTGTACGCCAGAACTAGCTGAACCTACTCAGTGAAGTCTTTTTCAGATTTTACTTGT 301
 QY 225 ProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIleAlaGlnLeuLeu 244
 Db 302 CCCCTAGTCGCAATACAGAAGCTGTGTGATGATCAAAAAGAGGACATTTGCTCAGCTTG 361
 QY 245 ThrArgGlnValLysGluProValArgPheThrGluSerIleGlyValMetGlnGluAla 264
 Db 362 ACGGTCAAGGTCAAGAACCCGTCGTTCTATGAAGATATTTGGGGTCATCGCAAGAAGCA 421
 QY 265 GlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGlyPheValLysLys 284
 Db 422 GGCATAACCACTATTTCGAGATTGGACCGGGGAAGTCTTGTCAAGTCTTGTATAAAA 481
 QY 285 IleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeuValAlaLeuLeu 304
 Db 482 ATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTAGTACACTTTTA 541
 QY 305 GluLys 306
 Db 542 GAAAAA 547

RESULT 14

AAT98542
 ID AAT98542 standard; DNA; 1196 BP.

XX AAT98542;

XX AC AAT98542;

XX DT 06-NOV-1998 (first entry)

XX DE DNA encoding a S. pneumoniae 3-oxoacyl reductase precursor protein.
 XX Streptococcus pneumoniae protein; genetic immunisation; antagonist;
 KW immunological response; inoculation; antibody production; inhibitor;
 KW T cell immune response; antimicrobial compound; bacterial adhesion;
 KW extracellular matrix protein; protein-mediated cell invasion; wound;
 KW pathogenesis; ss.

XX Streptococcus pneumoniae.

XX Key Location/Qualifiers
 XX CDS 586..1146
 XX /*tag= a

XX WO9743303-A1.

XX 20-NOV-1997.

XX 14-MAY-1997; 97WO-US07950.

XX 14-MAY-1996; 96US-0017670.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX (SMIK) SMITHKLINE BEECHAM PLC.

XX Black MT, Hodgson JE, Knowles DJC, Nicholas RO;

XX Stodola RK;

XX WPI; 1998-008993/01.

XX P-PSDB; AAW38474.

XX Novel Streptococcus pneumoniae proteins and related DNA - useful for
 diagnosing anti-microbial agents for treatment of bacterial
 infections

XX Claim 4; Page 90; 483pp; English.

XX This sequence encodes a Streptococcus pneumoniae protein that (based on
 CC homology with a Cuphea lanceolata protein) is a 3-oxoacyl reductase
 CC precursor protein, and represents a DNA sequence of the invention.
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
 CC invention can be used to identify compounds which interact with and
 CC inhibit or activate the activity of the proteins. Antagonists can be
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic
 CC immunisation. They can also be used to induce an immunological response
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
 CC of the encoding nucleic acids in a vector adequate to produce antibody
 CC and/or T cell immune responses to protect the animal from disease. The
 CC proteins can also be used to identify antimicrobial compounds which are
 CC capable of inhibiting their bioactivity. In particular the proteins of
 CC the invention can be used to prevent adhesion of bacteria to mammalian
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to
 CC block protein-mediated mammalian cell invasion, and to block the normal
 CC progression of pathogenesis in infections initiated other than by the
 CC implantation of in-dwelling devices or other surgical techniques.

XX SQ Sequence 1196 BP; 355 A; 223 C; 286 G; 332 T; 0 other;

Alignment Scores:

Pred. No.: 2,548-79 Length: 1196
 Score: 858.00 Matches: 180
 Percent Similarity: 98.37% Conservative: 1
 Best Local Similarity: 97.83% Mismatches: 2
 Query Match: 56.52% Indels: 2
 DB: 19 Gaps: 0

US-09-308-397-2 (1-306) x AAT98542 (1-1196)

QY 125 AlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGluGluAla 144
 Db 2 GCTGAACCTGGCAAGATGGTAGCAGTCTCTAATACGCCAGTAGAGTCAATGAAGAAGCC 61
 QY 145 CysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnThrAsnThrProAlaGln 164
 Db 62 TGTCAAAAGCTCTGGAACTTGGAGTGGTTACTCCAGCCAACTATAACACACCTGCACA 121
 QY 164 ntleValIleAlaGlyGluValValAlaValAlaAspArgAlaValGluLeuGlnGluAla 184
 Db 122 AATCGTCATTGCTGGAGAAGTGGTTGCAGTTCATCGAGCGGTTGAACCTTTTGAAGAAGC 181

QY 184 aGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeuLeuG 204
 DB 182 AGGTGCCAAACGCTTGATTCCTTAAGGTGTGAGGTCCCTTTACACCCGCTCTCTTGA 241
 QY 204 uProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPheThrCy 224
 DB 242 GCCTGCTAGCCAGAACTAGCTGAACCTCTAGCTCAGGTAAAGTTTTTCAGATTTTACTTG 301
 QY 224 sProLeuValGlyAsnThrGluAlaValMetGlnLysGluAspIleAlaGlnLeuLe 244
 DB 302 TCCCTTAGTCGGCAATACAGAGCTGCTGTGATGCAAAAGAGGACATTCCTCAGCTCTT 361
 QY 244 uThrArgGlnValLysGluProValArgPheThrGluSerIleGlyValMetGlnGluAl 264
 DB 362 GACCGCTCAGGTCAGGAACCGTTCGTTCTATGAAGTATTGGGGTCTATGCAAGAAGC 421
 QY 264 aGlyIleSerAsnPheIle-GluIleGlyProGlyLysValLeuSerGlyPheValLysL 284
 DB 422 AGGCATTAAGCAACTTTATTCGAGATTGGACCGGGGAAAGTCTTGTCAGGTTTGTAAAA 481
 QY 284 ySleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeuValAlaLeuL 304
 DB 482 AAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTAGTACACTTT 541
 QY 304 euGluLys 306
 DB 542 TAGAAAAA 549

RESULT 15

AAAX30801/c

ID AAX30801 standard; DNA; 1199 BP.

XX AC

XX AC

XX AC

XX 20-MAY-1999 (first entry)

XX DE Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:78.

XX KW Streptococcus pneumoniae strain 0100993; vaccine; immune response;

XX KW Streptococcus pneumoniae; pneumococcal; ss.

XX OS Streptococcus pneumoniae.

XX XX

XX FN W09737026-A1.

XX XX

XX PD 09-OCT-1997.

XX PF 01-APR-1997; 97WC-US05306.

XX XX

XX PR 22-AUG-1996; 96US-0025788.

XX PR 02-APR-1996; 96US-0014690.

XX XX

XX PA (SMIK) SMITHKLINE BEECHAM CORP.

XX PA (SMIK) SMITHKLINE BEECHAM PLC.

XX XX

XX PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;

XX PI Stodola RK;

XX XX

XX DR WPI; 1997-503111/46.

XX DR P-PSDB; AAY11212.

XX XX

XX PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in

XX PT vaccines, drug screening, etc

XX XX

XX PS Claim 5; Page 105-106; 354pp; English.

XX XX

XX CC AAX30724 to AAX30946 represent genomic DNA sequences isolated from

XX CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences

XX CC encode the novel proteins given in AAY1114 to AAY11367. The proteins,

XX CC isolated from Streptococcus pneumoniae, can be used in vaccines against

XX CC streptococcal infections and in assays for identifying compounds that

XX CC inhibit or activate the activity of the proteins. The antagonists can

CC be used to treat an individual having need to inhibit a bacterial
 CC protein. Vectors expressing the proteins can be used to induce a
 CC protective immune response in mammals.

XX Sequence 1199 BP; 334 A; 287 C; 223 G; 355 T; 0 other;
 SQ

Alignment Scores:

Pred. No.: 2,55e-79 Length: 1199
 Score: 858.00 Matches: 180
 Percent Similarity: 98.37% Conservative: 1
 Best Local Similarity: 97.83% Mismatches: 1
 Query Match: 56.52% Indels: 2
 DB: 18 Gaps: 0

US-09-308-397-2 (1-306) x AAX30801 (1-1199)

QY 125 AlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGluGluAla 144
 DB 1198 GCTGAACCTTGCACAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTCAATGAAGAAGCC 1139
 QY 145 CysGlnLysAlaSer-GluLeuGlyValValThrProAlaAsnTyrAsnThrProAlaG 164
 DB 1138 TGTCAAAAAGCTTCTGGAACCTTGGAGTGGTTACTCCAGCCAACTATAACACACCTGCACA 1079
 QY 164 nIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeuLeuGluAl 184
 DB 1078 AATCGTCATTCTCGAGAAGTGGTTGCAGTTGATCGAGCCGTTGAACCTTTTGAAGAAGC 1019
 QY 184 aGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeuLeuG 204
 DB 1018 AGGTGCCAAACGCTTGATTCCTCTTAAGGTGTGAGTCCCTTTACACCGCTCTCTTGA 959
 QY 204 uProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPheThrCy 224
 DB 958 GCCTGCTAGCCAGAACTAGCTGAACTCTAGCTCAGGTAAAGTTTTTCAGATTTTACTTG 899
 QY 224 sProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIleAlaGlnLeuLe 244
 DB 898 TCCCTTAGTCGGCAATACAGAAAGCTGCTGTGATGCAAAAAGAGGACATTTGCTCAGCTCTT 839
 QY 244 uThrArgGlnValLysGluProValArgPheThrGluSerIleGlyValMetGlnGluAl 264
 DB 838 GACCGCTCAGTCAAGGAACCCGTTCTGTTCTATGAAGTATTGGGGTCTATGCAAGAAGC 779
 QY 264 aGlyIleSerAsnPheIle-GluIleGlyProGlyLysValLeuSerGlyPheValLysL 284
 DB 778 AGGCATTAAGCAACTTTATTCGAGATTGGACCGGGGAAAGTCTTGTCAAGTTTGTAAAA 719
 QY 284 ySleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeuValAlaLeuL 304
 DB 718 AAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTAGTACACTTT 659
 QY 304 euGluLys 306
 DB 658 TAGAAAAA 651

Search completed: June 11, 2003, 20:37:37

Job time : 1188 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 11, 2003, 20:12:23 ; Search time 68 seconds
(without alignments)
1380.045 Million cell updates/sec

Title: US-09-308-397-2

Perfect score: 1518

Sequence: 1 MTKTAFLPAGGQAYLGMR.....QTAHLAHEVQASIVALLEK 306

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USPTO.spool/US09308397/runat 06062003 112349 28904/app query.fasta_1.455
-DB=Issued Patents.NA -QWMT=fastap -SUFFIX=rni -MINWATCH=0.1 -LOOPCI=0
-LOOPEXT=0 -UNITS=BITS -SPART=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFM=ptc -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09308397 -CGEN 1 1 36 -runat 06062003 112349 28904 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_THREADS=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents.NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	1518	100.0	19702	4	US-08-961-527-7
C 2	858	56.5	1196	4	US-08-858-207A-8
C 3	619	40.8	9686	4	US-09-221-017B-1003
C 4	547.5	36.1	927	1	US-08-789-609A-1
C 5	547.5	36.1	927	4	US-09-108-517-1
C 6	515.5	34.0	936	4	US-09-134-001C-1987
C 7	414.5	27.3	4403765	4	US-09-103-840A-2
C 8	414.5	27.3	4411529	4	US-09-103-840A-1
C 9	376.5	24.8	50937	4	US-09-428-517-1
C 10	374	24.6	68750	3	US-09-335-409-1
C 11	374	24.6	68750	4	US-09-568-102-1
C 12	374	24.6	68750	4	US-09-567-969-1

13	374	24.6	68750	4	US-09-568-480-1	Sequence 1, Appli
14	374	24.6	68750	4	US-09-568-486-1	Sequence 1, Appli
15	374	24.6	68750	4	US-09-568-472-1	Sequence 1, Appli
16	374	24.6	68750	4	US-09-567-899-1	Sequence 1, Appli
17	372.5	24.5	15872	4	US-09-105-537-1	Sequence 1, Appli
C 18	372	24.5	4403765	4	US-09-103-840A-2	Sequence 2, Appli
C 19	372	24.5	4411529	4	US-09-103-840A-1	Sequence 1, Appli
C 20	369	24.3	71989	4	US-09-443-501A-2	Sequence 2, Appli
C 21	367.5	24.2	43280	2	US-08-804-227C-1	Sequence 1, Appli
C 22	349.5	23.0	33529	4	US-09-144-085-3	Sequence 3, Appli
C 23	348.5	23.0	8268	1	US-08-375-709-10	Sequence 10, Appli
C 24	348.5	23.0	8268	1	US-08-752-929-10	Sequence 10, Appli
C 25	348.5	23.0	37895	1	US-08-375-709-1	Sequence 1, Appli
C 26	348.5	23.0	37895	1	US-08-752-929-1	Sequence 1, Appli
C 27	348.5	22.7	37895	3	US-09-090-793-1	Sequence 1, Appli
C 28	344.5	22.7	80161	3	US-09-036-987A-1	Sequence 1, Appli
C 29	344.5	22.7	80161	4	US-09-370-700-1	Sequence 1, Appli
C 30	343.5	22.6	19227	3	US-09-090-793-13	Sequence 13, Appli
C 31	343.5	22.6	40138	3	US-09-090-793-12	Sequence 12, Appli
C 32	334	22.0	1030	3	US-08-858-003-2	Sequence 2, Appli
C 33	334	22.0	1030	3	US-09-078-166-2	Sequence 2, Appli
C 34	334	22.0	1030	4	US-08-997-467-2	Sequence 2, Appli
C 35	329.5	21.7	49377	1	US-08-764-233A-1	Sequence 1, Appli
C 36	324	21.3	13842	4	US-09-105-537-30	Sequence 30, Appli
C 37	324	21.3	36778	4	US-09-105-537-5	Sequence 5, Appli
C 38	324	21.3	38506	3	US-09-320-878-19	Sequence 19, Appli
C 39	316.5	20.8	28958	1	US-08-258-261B-6	Sequence 6, Appli
C 40	316.5	20.8	28958	1	US-08-456-837-6	Sequence 6, Appli
C 41	316.5	20.8	28958	1	US-08-457-342-6	Sequence 6, Appli
C 42	316.5	20.8	28958	1	US-08-457-648A-6	Sequence 6, Appli
C 43	316.5	20.8	28958	1	US-08-458-076A-6	Sequence 6, Appli
C 44	316.5	20.8	28958	1	US-08-764-233A-4	Sequence 4, Appli
C 45	316.5	20.8	28958	1	US-08-457-335A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-08-961-527-7/c
; Sequence 7, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,527

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

```
;
; LENGTH: 19702 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-7
Alignment Scores:
Pred. No.: 2 78e-166 Length: 19702
Score: 158.00 Matches: 306
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-308-397-2 (1-306) x US-08-961-527-7 (1-19702)
QY 1 MetThrIysThrAlaPheLeuPheAlaGlyGlnGlyValaGlnTyrLeuGlyMetGlyArg 20
Db 17149 ATGACTAAACAGCGCTTTTATTGCTGTCAGAGTCCCGAGTATCTAGGATGGACGG 17090
QY 21 AspPheTyrAspGlnTyrProIleValIysGluThrIleAspArgAlaSerGlnValLeu 40
Db 17089 GATTTCATGATCAGTATCCGATGTTTAAAGAACGATTGATCGAGCGAGCGGTGCTC 17030
QY 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluAspIysLeuAsnGlnThrArgTyr 60
Db 17029 GGTATGATTGCGTTATCTCATCGATACGGAAGAGACAACTCAATCAGACCGCTAT 16970
QY 61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGlnIysGly 80
Db 16969 AGCAACACCGCATTTAGGAGTTCGGTTGCTATCTACCGTTTATGCAAGAAAGGCG 16910
QY 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
Db 16909 TATCAGCCTCATATGTTGCTGTTGCTCTTGGAGATACTCTGCTTGGTGGCAAGT 16850
QY 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaIysArgGlyAlaTyrMetGlu 120
Db 16949 GGGCGCTTGATTGTAAGATGGGTTGCTGCTGAGTAAAGCGTGGACCTATATGGA 16790
QY 121 GluAlaAlaProAlaAspSerGlyIysMetValAlaValIleAsnThrProValGluVal 140
Db 16789 GAAGCGCTCTCTGCTGCTGCAAGATGGTAGCAGTTCTCATACGCCAGTAGAGTC 16730
QY 141 IleGluAlaCysGlnIysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
Db 16729 ATTGAAGAGCGCTGCAAAAGCTTCTGAACCTGGAGTGGTTACTCCAGCCAACTATA 16670
QY 161 ThrProAlaGlnIleValIleAlaGlyGluValAlaValAlaAspArgAlaValGluLeu 180
Db 16669 ACACCTGCACAAATCGTTCATTGCTGGAGAGTGGTTGCGAGTTGATCGAGCGTTGA 16610
QY 181 LeuGlnGluAlaGlyAlaIysArgLeuIleProLeuIysValSerGlyProPheHisThr 200
Db 16609 TTGCAAGAGAGCGTGCACAAAGCTTGAATCTCTTAAGTGTGAGTCCCCTTCACACC 16550
QY 201 AlaLeuLeuClubProAlaSerGlnIysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
Db 16549 GCTCTCTTGAACCTGCTAGCCAGAACTAGCTGAAACTCTAGCTAGGTAAGTTTTC 16490
QY 221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnIysGluAspIle 240
Db 16489 GATTTTACTTGTCCTAGTCGCAATACAGAGCTGCTGTGATGCAAAAGAGACATT 16430
QY 241 AlaGlnLeuLeuThrArgGlnValIysGluProValArgPheTyrGluSerIleGlyVal 260
Db 16429 GCTCAGCTCTGACGGCTCAGGTCAGGAAACCGGTTGCTTCTATGAAGATTATCGGTC 16370
QY 261 MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyIysValLeuSerGly 280
Db 16369 ATCAAGAGACGCGATAGCACTTATCGAGATTGGACCGGGAAGCTCTTGTGAGGT 16310
QY 281 PheValIysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
```

```
Db 16309 TTTGTTAAAAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAACGAGTTTA 16250
QY 301 ValAlaLeuLeuGluIys 306
Db 16249 GTAGCACCTTTTAGAAAAA 16232

RESULT 2
US-08-858-207A-8
; Sequence 8, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Robert
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: NO. 6348328e1 Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1196 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-858-207A-8
Alignment Scores:
Pred. No.: 5 88e-91 Length: 1196
Score: 858.00 Matches: 180
Percent Similarity: 98.37% Conservative: 1
Best Local Similarity: 97.83% Mismatches: 1
Query Match: 56.52% Indels: 2
DB: 4 Gaps: 0

US-09-308-397-2 (1-306) x US-08-858-207A-8 (1-1196)
QY 125 AlaAspSerGlyIysMetValAlaValLeuAsnThrProValGluValIleGluGluAla 144
Db 2 GCTGAACCTTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGTTCATTGAAGAGCC 61
QY 145 CysGlnIysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThrProAlaG 164
Db 62 TGTCAAAAAGCTTCTGGAACCTTGGAGTGTGTCTCCAGCCCACTATAACACACTGAC 121
QY 164 nileValIleAlaGlyGluValValAlaValAspArgAlaValGluLeuGlnGluAl 184
```



```

Db 122 AATCGTCATGCTGGAGAGTGGTTGCAGTTGATCGAGCGTTGAACCTTTGCAAGAGC 181
Qy 184 aGlyAlaLysArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 204
Db 182 AGGTGCAACAGCTTCTTAAAGTGTTCAGTCCCTTTCACCGCTCTCTTGA 241
Qy 204 uproAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPheThrCy 224
Db 242 GCTGCTAGCCGAACTAGCTGAACTCTGCTCAGGTAAGTTTTTTCAGATTTTACATG 301
Qy 224 spLeuValGlyAsnThrGluAlaValMetGlnLysGluAspLeuAlaGlnLeuLe 244
Db 302 TCCCTAGTCGCAATACAGAGCTGCTGTGATGCAAAAGAGGACATTCCTCAGCTCTT 361
Qy 244 uThrArgGlnValLysGluProValArgPheThrGlnSerLeuLeuGlnValMetGlnL 264
Db 362 GACGGCTCAGGTCAAGAAACCGCTTCTTCTATGAAAGTATTTGGGTCTATGCAAGAGC 421
Qy 264 aGlyLeuSerAsnPheLeu-GluLeuGlyProGlyLysValLeuSerGlyPheValLysL 284
Db 422 AGGCATACCACTTATTCAGATTTGGACCGGGGAAAGTCTTTGTCAGGTTTGTAAAA 481
Qy 284 yIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeuValAlaLeuL 304
Db 482 AAATTGATCAAACTGCTCCTTAGCTCATGTGCAAGATCAAGCGAGTTTAGTAGCATT 541
Qy 304 euGluLys 306
Db 542 TAGAAAAA 549

```

RESULT 3

```

US-09-221-017B-1003/c
; Sequence 1003, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Roes, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIORITY INFORMATION:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCES/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600

```

```

TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 1003:
SEQUENCE CHARACTERISTICS:
LENGTH: 9686 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...9686
US-09-221-017B-1003
Alignment Scores:
Pred. No.: 1-27e-61 Length: 9686
Score: 619.00 Matches: 132
Percent Similarity: 65.72% Conservative: 54
Best Local Similarity: 46.64% Mismatches: 93
Query Match: 40.78% Indels: 4
DB: Gaps: 4
US-09-308-397-2 (1-306) x US-09-221-017B-1003 (1-9686)
Qy 5 AlapheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyrAsp 24
Db 4534 GCATTTGTATTCCTCCCGTCCAGGAGCAGCATGTTCTAGTAATGGAAAAAGACCTGTACGAA 4475
Qy 25 GlnTyr-ProfileValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyrAspLeu 44
Db 4474 CAGAACCCCGAGGCCAAAGCATATTTGAAAGCCCAATGAGATATTAGCTTTCGCATT 4415
Qy 45 Arg---TyrLeuIleAspThrGluGluAsnLysLeuAsnGlnThrArgTyrThrGlnPro 63
Db 4414 ACGGACATCTCTTCAACGCGCAGCAGAGAGCTGAAACCAACCAAGTACCCAGCG 4355
Qy 64 AlaIleLeuAlaThrSerValAlaIleTyrArgLeuGlnGluLysGlyTyrGlnPro 83
Db 4354 GCGATCTCTCCACTCGGTGATTCTGGCCAGACGATGGCGAA--GACTTCCCTCCC 4298
Qy 84 AspMetValAlaGlyLeuSerLeuGlyTyrSerAlaLeuValAlaSerGlyAlaLeu 103
Db 4297 GATATGGTGCAGGCCATTCGCTCGAGAGTCTTTCGCAATTTGTCAGCAGGTGCATG 4238
Qy 104 AspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluGluAlaA 123
Db 4237 ACCTTCGAAGATGTTCTCTGCTCTGATCGAAACGTGCCATGGCCATGCAGAAAGCATGC 4178
Qy 124 ProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGluGlu 143
Db 4177 GAAGTACGCCCTCGACTATGGCGGTCTGACTGGGTCTGCTGATGAAAGGTGGAAGAA 4118
Qy 144 AlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThrProAla 163
Db 4117 ATCTGTGCCCAAGTACCGAT---GAAGTTGTGCTACCGGCCCAACTACAACTGCCCGGGA 4061
Qy 164 GlnIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeuGluGlnGlu 183
Db 4060 CAGATGTTATATCCGTTCCGTTAGAGGAGTGGACAGAGCTTTCGAACTGCTCAAGGAA 4001
Qy 184 AlaGlyAlaLysArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 203
Db 4000 GCCGGAGCAAGCGTCTCTTAAGTCTGCTGTCCGTGGAGAGCTTCATTCCTCTCTGATG 3941
Qy 204 GluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPheThr 223
Db 3940 GAACCGGCTCGTGAAGAGTGGCAAGCCATCAGGAGACTCATGATCAGCCGATC 3881
Qy 224 CysProLeuValGlyAsn---ThrGluAlaAlaValMetGlnLysGluAspIleAlaGln 242

```

DB 3880 TGCCCCATCTATCAGACGTTAACCGCTTCTCCGCTCCCGATCCGCGCAGAGATTAAG 3821
 QY 243 LeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGln 262
 DB 3820 AACCTGATAGCACAGCTGACGGCACCCTGCTGGACACAGAGTGCTGGAATATGACA 3761
 QY 263 GluAlaGlyLeuSerAsnPheIleGluIleGlyProGlyLysValLeuSerGlyPheVal 282
 DB 3760 GCCGATGCTGCCACCACTTTATGGAGTTAGGCCCGGAAATGATTGCAGGGTTGGTG 3701
 QY 283 LysLysIle 285
 DB 3700 AAGAAAT 3692

RESULT 4
 US-08-789-609A-1
 ; Sequence 1, Application US/08789609A
 ; Patent No. 5827689
 ; GENERAL INFORMATION:
 ; APPLICANT: Gentry, Daniel
 ; APPLICANT: Lonsdale, John
 ; APPLICANT: Pearson, Stewart
 ; APPLICANT: Payne, David
 ; TITLE OF INVENTION: No. 5827689e1 PabD
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SmithKline Beecham Corporation
 ; STREET: 709 Swedeland Road
 ; CITY: King of Prussia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19406-0939
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/789,609A
 ; FILING DATE: 24-JAN-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/030685
 ; FILING DATE: 13-NOV-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gimmi, Edward R
 ; REGISTRATION NUMBER: 38,891
 ; REFERENCE/DOCKET NUMBER: GMS0004
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 610-270-4478
 ; TELEFAX: 610-270-5090
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 927 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; US-08-789-609A-1

Alignment Scores:
 Pred. No.: 8,15e-55 Length: 927
 Score: 547.50 Matches: 121
 Percent Similarity: 58.72% Conservative: 54
 Best Local Similarity: 40.60% Mismatches: 118
 Query Match: 36.07% Indels: 5
 DB: 1 Gaps: 4

US-09-308-397-2 (1-306) x US-08-789-609A-1 (1-927)

QY 1 MetThrLysThrAlaPheLeuAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20

DB 1 ATGAGTAAACAGCAATTTATTTTCGGGCAACAGGTGCCCAAAAGTTGGTATGGCAAA 60
 QY 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
 DB 61 GATTGGTTTAAATAACATGATCAACCACTGAAATTTTAACTTCAGCAGCAACACGTTA 120
 QY 41 GlyTyrAsp---LeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArg 59
 DB 121 GACTTTGATATTTTAGACAACAATTTACTCATGAAGAAGGTAAATTTGGGTGAAACTGAA 180
 QY 60 TyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGluGlnGluLys 79
 DB 181 AACCGCAACCCAGCTTTATTGACCATAGTTTCGGCATTTATTAGCGGCTA-----AAA 234
 QY 80 GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAla 99
 DB 235 ATTTTGAATCCTGATTTTACTATGGGCATAGTTTAGTGGAATATTCAAGTTTAGTTGCA 294
 QY 100 SerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMet 119
 DB 295 GCTGACGTATTATCATTTGAAGATGCAATTTAAATTTGTAGAAAACGTGGTCAATTAATG 354
 QY 120 GluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGlu 139
 DB 355 GCGCAAGCATTTCTACTGTTGTAGGAAGCATGCTGCAGTATTGGGATTAGATTTTCAT 414
 QY 140 ValIleGluGluAlaCysGlnLys---AlaSerGluLeuGlyValValThrProAlaAsn 158
 DB 415 AAGTCGATGAATTTTGTAGTCATTTATCATCTCATGACAAATAATTTGAACCCAGCAAC 474
 QY 159 TyrAsnThrProAlaGlnIleValIleAlaGlyLeuValAlaValAlaAspArgAlaVal 178
 DB 475 ATTAATTCGCCAGGTCAAATTTGTTTCAGTCCAAAGCTTTAATTTAGTACGTAGTA 534
 QY 179 GluLeuLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPhe 198
 DB 535 GAAAAAGGTAAATCATTTAGGTGCAAAACGTCATGCTTTAGCAGTATCTGGACCATTC 594
 QY 199 HisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSer 218
 DB 595 CATTCATCGCTAATGAAGATGATTGAAGAGATTTTCAAGTTACATTAATCAATTTGAA 654
 QY 219 PheSerAspPheThrCysProLeuValGlyAsnThrGluAla---AlaValMetGlnLys 237
 DB 655 TGGCGTGTATGCTAAAGTTTCTGTTAGTTCATAAATGCAAAAGGTGAAAGTGCACAAA 714
 QY 238 GluAspIleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSer 257
 DB 715 GAAGTAATTAATCTAATATGTTGTCACGAATTTATTTCCACGAGTACATTCATTACTCA 774
 QY 258 IleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysVal 277
 DB 775 ACAGATGGCTAATAGACCAAGGTGTTGATCATTTTATTGAAATTTGCTCTGGAAGATT 834
 QY 278 LeuSerGlyPheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGlu 295
 DB 835 TTATCTGCTTATTAATAAAAAATAATAGAGATGTTTAAGTTAATCAATCAATTCAA 888

RESULT 5
 US-09-108-517-1
 ; Sequence 1, Application US/09108517
 ; Patent No. 6258934
 ; GENERAL INFORMATION:
 ; APPLICANT: Gentry, Daniel
 ; APPLICANT: Lonsdale, John
 ; APPLICANT: Pearson, Stewart
 ; APPLICANT: Payne, David
 ; TITLE OF INVENTION: No. 6258934e1 PabD
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SmithKline Beecham Corporation
 ; STREET: 709 Swedeland Road

```

; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/108,517
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/789,609
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: GM50004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 927 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-09-108-517-1

Alignment Scores:
Pred. No.: 8,15e-55 Length: 927
Score: 547.50 Matches: 121
Percent Similarity: 58.7% Conservative: 54
Best Local Similarity: 40.60% Mismatches: 118
Query Match: 36.07% Indels: 5
DB: 4 Gaps: 4

US-09-308-397-2 (1-306) x US-09-108-517-1 (1-927)
QY 1 MetThrLysThrAlaPheLeuPheAlaGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 ATGAGTAAACAGCAATTATTTCCGGACAGAGTGCCTGAGGAGGTTGGTATGGCAAA 60
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 GATTTGTTTAAACAATGATCAAGCACTGAAATTTTAACTTCAGCAGCAAGACGTTA 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 41 GlyTyrAsp---LeuArgTyrIleuIleAspThrGluGluAspLysLeuAsnGlnThrArg 59
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GACTTTGATATTTTATAGACAACTGTTTACTGATGAAGAGGTTAAATGGGTGAACCTGAA 180
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 60 TyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLys 79
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 AACAGCCACACAGCTTTTATGAGCATAGTTCGGCATTTATAGCAGCGCTA-----AAA 234
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 80 GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAla 99
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 235 ATTTTGAATCTCTGATTTTACTATGGGCATAGTTAGGTGAATATTCAAGTTAGTTGCA 294
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 100 SerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMet 119
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 295 GCTGAGCTATTATCATTTTGAAGATGCACTTTAAATTTGTTAGAAAACGTGGTCAATTAATG 354
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 120 GluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGlu 139
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 355 GCGCAAGCATTTCTCTACTGGGTAGGAAGCATGGCTGCGCAGTATTGGGATTAGATTTTGTAT 414
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 140 ValIleGluGluAlaCysGlnLys---AlaSerGluLeuGlyValValThrProAlaAsn 158
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 415 AAAGTCGATGAAATTTGTAGTCATTATCATCTGATGACAAAATAATTAATGAACAGCAAAAC 474
QY 159 TyrAsnThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaVal 178
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 475 ATTAATTGCCAGGTCAAATTTGTTTCCAGGTCACAAAAGCTTAAATTTGATGAGCTAGTA 534
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 179 GluLeuLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPhe 198
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 535 GAAAAAGGTAATCATTTAGGTGCAAAACGTGTCATGCTCTTAGCAGTATCTTGACCATTC 594
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 199 HisThrAlaLeuLeuProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSer 218
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 595 CATTCATCGCTAATGAAGTGTGAAGAGATTTTTCAGTTACATTATTAATCAATTTGAA 654
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 219 PheSerAspPheThrCysProLeuValGlyAsnThrGluAla---AlaValMetGlnLys 237
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 655 TGGCGTGATGCTAAGTTTCTGCTAGTTCAAAATGTAATGCGCAAGGTGAAACTGACAAA 714
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 238 GluAspIleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSer 257
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 715 GAAGTAAATTAATCTAATATGTCACAGCAATATATTCACCACTACAAATTCATTAAC 774
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 258 IleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysVal 277
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 775 ACAGAATGGCTAATAGACCAAGGTGTGATCATTTTATTCAAATTTGCTCTCGAAAAGTT 834
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 278 LeuSerGlyPheValLysIleAspGlnThrAlaHisLeuAlaHisValGlu 295
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 835 TTAATCTGGCTTAATTAATAAATAATAGAGATGTTAAGTTAACATCAATTCAA 888
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
US-09-134-001C-1987
; Sequence 1987, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1987
; LENGTH: 936
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-1987

Alignment Scores:
Pred. No.: 4,57e-51 Length: 936
Score: 515.50 Matches: 116
Percent Similarity: 58.42% Conservative: 61
Best Local Similarity: 38.28% Mismatches: 111
Query Match: 33.96% Indels: 15
DB: 4 Gaps: 5

US-09-308-397-2 (1-306) x US-09-134-001C-1987 (1-936)
QY 1 MetThrLysThrAlaPheLeuPheAlaGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10 ATGGGTAAATAGCAGTGTCTTCCCTGGCCAAAGGTGCACAAAAGATGGGATGCTAGT 69
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 70 GACTTATATAATGAAGAGGAAAGCTCGACTCGAGTATTAAATTTAGCTCAAGACAGTA 129
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAsp---LysLeuAsnGlnThrArg 59
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 GACTTTGATTTGCTAGAAAACGATGTTTCTGATAATGATGGGAATTTAGGAGAAACAGAA 189
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```
QY 60 TyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLys 79
Db 190 AATACACAACTGACATTAATACACATAGTGTTCGGTACTAGAAATCATTAATAATCAT 246
QY 80 GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGluTyrSerAlaLeuValAla 99
Db 247 ---TTAAAGTAGATTAATACATATGGACATAGCTTAGAGAAATATTCAGCTTAGTTCT 303
QY 100 SerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMet 119
Db 304 AGTGATGATTAATCTTTTGAAGATCGCGTGTAGAAATTTGGCTAAACGTGCCCACTAATG 363
QY 120 GluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuLeuThrProValGlu 139
Db 364 GCTCAAGCGTTCTTAACGGTGTGGAGGTATGGCAGCATATTAGTTTGGATTATGAT 423
QY 140 ValIleGluGluAlaCysGlnLysAlaSer---GluLeuGlyValValThrProAlaAsn 158
Db 424 GATGTTGATAAGATATGTCAAACGTTATCTACAAAGAACAGTAAATTGAACCTGCTAAT 483
QY 159 TyrAsnThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaVal 178
Db 484 ATTAACCTCACAGTCAATTCGTGTCTGGACATAAATCTTTAATTGATGAATTAGTA 543
QY 179 GluLeuLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPhe 198
Db 544 GAAAGGGCAAGAACTTGGTGTAAACGTTGTCTTCCATTAGCTGTTCCGGTCTCTTTT 603
QY 199 HisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSer 218
Db 604 CATCTCTCAATGATGAAGTATTGAAGAGATTTTCTAATTAATTAATCAATTAATGA 663
QY 219 PheSerAspPheThrCysProLeuVal-----GlyAsnThrGluAla 232
Db 664 TGCATAATGCTAATTAATCCAGTGTTCAGAACTGTTAATCAAAAGGAGAAACCGATGCT 723
QY 233 AlaValMetGlnLysGluAspIleAlaGlnLeuLeuThrArgGlnValLysGluProVal 252
Db 724 GAAGTAATTAACCAAT-----ATGGTTAAACAAATATATTCACCTGTT 768
QY 253 ArgPheTyrGluSerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIle 272
Db 769 CAATTTATCAATCAACGGAGTGTGTGATTAATCAAGTGTGATCACTTTATTGAAT 828
QY 273 GlyProGlyLysValLeuSerGlyPheValLysLysIleAspGlnThrAlaHisLeuAla 292
Db 829 GGACCGGGAAGATTAATCTGGCTTTATCAAAAAATAAATCGAGATGTAAAAATCACT 888
QY 293 HisValGlu 295
Db 889 TCATTTCAA 897

RESULT 7
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis

;
FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 1,188-33 Length: 4403765
Score: 414.50 Matches: 115
Percent Similarity: 53.47% Conservative: 39
Best Local Similarity: 39.93% Mismatches: 117
Query Match: 27.31% Indels: 17
DB: 4 Gaps: 8

US-09-308-397-2 (1-306) x US-09-103-840A-2 (1-4403765)
QY 3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22
Db 489058 AAGACGGCTTTTGTGTTTGGCGGTTCAGGCTCGAGCTGGCTGGGTATGGTAGCAGCTT 489117
QY 23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
Db 489118 TATGCTGCTTACCGGTTTTCGGAGCCCTCGATGCTGTGGTGGACGAGTTGGACCGG 489177
QY 43 AspLeuArgTyr---LeuIleAspThr-----GluGluAspLysLeuAsnGln 57
Db 489178 CACCTGCGGTATCGCTCGCGATGTATCTGGGGCAGCAGCAAGATCTGTGAATACC 489237
QY 58 ThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGln 77
Db 489238 ACCGAATTCGCCAGCGCGCTGTTTCGGTGGAGTGGCGCTGTATCGGCTGCTCATG 489297
QY 78 GluLysGlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeu 97
Db 489298 TCGTGGGGGTGCGCGCGGTTTGTGTGCTGCTCATTCGTTGGCGAGTGTGGCGCGCGC 489357
QY 98 ValAlaSerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAla 117
Db 489358 CACGTCGCGGGCGCGCTGTTTCCGCGATCGCGCATGTGGTGGCGCGCGCTGGACGG 489417
QY 118 TyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrPro 137
Db 489418 TTGATG---CAGCGGTTCCCGCC---GGCGCGCCCATGTTTGGCGTGCAGCGCGTGAA 489471
QY 138 ValGluValIleGluGluAlaCysGlnLysAlaSerGluLeuGly---ValValThrPro 156
Db 489472 GACGAGGTA-----CCGCGCATGCTGGGGCAGCATGTGAGCATC 489510
QY 157 AlaAsnTyrAsnThrProAlaGlnIleValIleAlaGlyGluValAlaValAspArg 176
Db 489511 CCGCGGTCATGTCGCGCTTCGGTGGTGTATCTCTGTTGCCACGATGCGGTGAGCGCG 489570
QY 177 AlaValGluLeuLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGly 196
Db 489571 ATCGCTGATCGCTGCGCGCGCAGCGC---CGTCGGGTCCACCGGTTGGCGGTCTCCCAT 489627
QY 197 ProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGln 216
Db 489628 GCCTTTCACTCGGCGTTGATGGACCGCATGATCCGTGATTCAGTTCACAGCGGTTCGCCGAA 489687
QY 217 ValSerPheSerAspPheThrCysProLeuValGlyAsn---ThrGluAlaAlaValMet 235
Db 489688 CTGTCTGGGCTTGGCCCGCATATCCCGTCAATTTCCCAATGTGACCGGCGATTTGGTGCC 489747
QY 236 GlnLysGluAspIleAlaGlnLeuThrArgGlnValLysGluProValArgPheTyr 255
Db 489748 GACGACTTCGCGCTCAGCTGATTACTGGGCCCGCATATCCGGCGGTGGTGGCGGTTTGGC 489807
QY 256 GluSerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGly 275
Db 489808 GACAGTGTCTGATGTCCTGCTGCGCGCTGCGCGCTGCTGCTTTCATCGAAGTCGCGCGCT 489867
QY 276 LysValLeuSerGlyPheValLys 283
```

```

Db 489868 GCGCGCTTGACGTGCTTGATCGAG 489891
RESULT 8
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
Alignment Scores:
Pred. No.: 1.19e-33 Length: 4411529
Score: 414.50 Matches: 115
Percent Similarity: 53.47% Conservative: 39
Best Local Similarity: 39.93% Mismatches: 117
Query Match: 27.31% Indels: 17
DB: 4 Gaps: 8
US-09-308-397-2 (1-306) x US-09-103-840A-1 (1-4411529)
Qy 3 LysThrAlaPheLeuPheAlaGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22
Db 487616 AGACGGCTTTGTGTTCCCGGTGAGGCTCGCATGGCTGGGTATGGGTAGCAGGTT 487675
Qy 23 TyrAspGlnTyrProLeuValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
Db 487676 TATGTCGCTACCGGCTTTCCCGAGGCCCTCGATGCTGTGTGGTGGACGAGTTGACCGG 487735
Qy 43 AspLeuArgTyr---LeuIleAspThr-----GluLysAspLysLeuAsnGln 57
Db 487736 CACCTCGGTATCCGTCGCGCATGTGATCGGGGCGACGACCAAGATCTGTGAATACC 487795
Qy 58 ThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGln 77
Db 487796 ACCGAATCCCGACCGCGCTGTTGGGTGGAGTGGCGCTGTATCGGCTGCTCATG 487855
Qy 78 GluLysGlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeu 97
Db 487856 TCGTGGGGGTGGCGCGGTTTGTGCTGGTTCATTCGTTGGGAGTGGCGCGCG 487915
Qy 98 ValAlaSerGlyValAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAla 117
Db 487916 CAGTCGCGGCGGCTGTTGTTGCGGATGCGGCGATGCTGTGTGGCGCGCGGAGCGG 487975
Qy 118 TyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrPro 137
Db 487976 TTGATG---CAGCGGTGGCGGCC---GGCGGCGCATGTTTGGCGTGCAGGCCCGTGAA 488029
Qy 138 ValGluValIleGluLysAlaSerGlnLysAlaSerGluLeuGly---ValValThrPro 156
Db 488030 GACGAGGTA-----GGCGGATGCTGGGCGACGATGTGAGCATC 488068
Qy 157 AlaAsnTyrAsnThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArg 176
Db 488069 GCGGCGGTCAATGGTCCGCTGTTGGTGGTATCTGTGTGGCCACGATGCGGTGAGCGG 488128
Qy 177 AlaValGluLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGly 196

```

```

Db 488129 ATCGCTGATCGGTGCGGCGCAGGCG---CGTCGGGTACACCGTTGCGGCTCGCAT 488185
Qy 197 ProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGln 216
Db 488186 GCCTTTCACTCGCGGTGATGGAGCGCATGATCGCTCAGTTACAGCCGTTGCGGCGAA 488245
Qy 217 ValSerPheSerAspPheThrCysProLeuValGlyAsn---ThrGluAlaAlaValMet 235
Db 488246 CTGTCTGTGGGCTTGGCCAGCATCCCGGTCAITTCATGTGACCGGCGAGTTGGTGCC 488305
Qy 236 GlnLysGluAspIleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyr 255
Db 488306 GACGACTTCGCTCAGCTGATTACTGGCGCGCATATCCCGGCGGTGGTGGCTTGGC 488365
Qy 256 GluSerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGly 275
Db 488366 GACAGTGTTCGTAGTCCCATCGCGGTCGCGAGTCGTTTCATCGAAGTCGGGCGCGGT 488425
Qy 276 LysValLeuSerGlyPheValLys 283
Db 488426 GCGCGCTTGACGTGCTGTGATCGAG 488449
RESULT 9
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1
Alignment Scores:
Pred. No.: 3.65e-32 Length: 50937
Score: 376.50 Matches: 99
Percent Similarity: 50.67% Conservative: 52
Best Local Similarity: 33.22% Mismatches: 122
Query Match: 24.60% Indels: 25
DB: 4 Gaps: 5
US-09-308-397-2 (1-306) x US-09-428-517-1 (1-50937)
Qy 5 AlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyrAsp 24
Db 7458 GCCTTCCTCTCTCCGGCAGGCGCAGCAGCCGCGCATGGCAAGCGACTCCACCG 7517
Qy 25 GlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyrAspLeu 44
Db 7518 GTGTTCCTCCGGCTTCGGGAGCGCCCTGGAGAGGTCTGCGCCGAACTCGACCCACCTC 7577
Qy 45 ArgTyrLeuIleAspThrGluGluAspLys----- 54
Db 7578 GGCGAGCTCTCGCGCCGAGCGCGCCCGCCCTGCGGACGCTGATGTTGCGCGAGCGG 7637
Qy 55 -----LeuAsnGlnThrArgTyrThrGlnProAlaIleLeuAla 67

```

Db 7638 GGCACGGCGCACAGCGCCCTCTCTCCAGAGCCACCTACACCCAGCGCCCTCTTGGCC 7697
Qy 68 ThrSerValAlaIleTyrArgLeuLeuGlnGlnLysGlyTyrGlnProAspMetValAla 87
Db 7698 CTGGAAACCGCCCTCTCTCCGCTCTCTGTCTCAGTGGGCGCTGAAACCGACCCCTCGCA 7757
Qy 88 GlyLeuSerLeuGlyGlyTyrSerAlaLeuValAlaSerGlyAlaLeuAspPheGluAsp 107
Db 7758 GGCACCTCCGTCGGCGAGATCGCGCGCGCCACACAGCAGCATCTCTCGACCTGTCCGAC 7817
Qy 108 AlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluGlnAlaAlaProAlaAspSer 127
Db 7818 GCGCGCAACTCGTGGCCACCGCGCGGTGTGTGCTTCCCTG-----CCCGCGCG 7871
Qy 128 GlyLysMetValAlaValLeuAsnThrProValGlnValleGluGlnAlaCysGlnLys 147
Db 7872 GCGCTCATGCTCTCGGTCCAGGACCCCGAGTCCGAGGTCCGCCCTCTGTCTCGCGCGT 7931
Qy 148 AlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThrProAlaGlnIleValle 167
Db 7932 GAGGCCACGTCGGCGT-----GCCGCCGTGAACGCCCGCCAGCGGTGTGTGTG 7982
Qy 168 AlaGlyGluValValAlaValAspArgAlaValGluLeuLeuGlnGlnAlaLys 187
Db 7983 TCCGCGAGCGCGCCACGTCGCGCGCATCGAACAGATCTCTCCGGACAGCGCGCGCAA 8042
Qy 188 ArgLeuIleProLeuLysValSerClyProPheHisThrAlaLeuLeuGluProAlaSer 207
Db 8043 AGCCGGTAC---CTGCGCGTCAAGCAGCCTCTCCACCTCCCGCTCATGGAACCGGTCTG 8099
Qy 208 GlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPheThrCysProLeuVal 227
Db 8100 GAGGAGTTCGCGAAGCGTCCGCGCTGACCTTCCGGGACCGACCGACCCCTGTC 8159
Qy 228 GlyAsnThrGluAlaAlaValMetGlnLysGluAspIleAla-----GlnLeuLeuThr 245
Db 8160 TCCAACTCACCGCGCACCATGTCAGACCGGACCATGGCCACCGCCGCTACTGGGTC 8219
Qy 246 ArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGlnGlnAlaGly 265
Db 8220 CGGCACGTCGCGAAGCGTCCGCTTCGGCGACCGCATCCGGGACATCGGGAACTGGGC 8279
Qy 266 IleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGlyPheValLys 283
Db 8280 ACCGCGACCTTCTGGAAGTCGCGCGACGCGCTCTCTCCGCCATGGCGCGC 8333

RESULT 10
US-09-335-409-1

; Sequence 1, Application US/09335409

; Patent No. 6121029

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/335,409

; CURRENT FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 68750

; TYPE: DNA

; ORGANISM: Sorangium cellulosum

US-09-335-409-1

Alignment Scores:

Pred. No.: 1.13e-31

Score: 374.00

Length: 68750

Matches: 101

Percent Similarity: 47.93% Conservative: 38
Best Local Similarity: 34.83% Mismatches: 135
Query Match: 24.64% Indels: 16
DB: 3 Gaps: 6
US-09-308-397-2 (1-306) x US-09-335-409-1 (1-68750)

Qy 3 LysThrAlaPheLeuPheAlaGlyGlnIleTyrLeuGlyMetGlyArgAspPhe 22
Db 56591 AAGCTCGCTTCTCTCTCCGCGCGAGCGCGCGAGTTCGCGCGCATCGGGCTGGTGTG 56650
Qy 23 TyrAspGlnTyrProIleValLysGlnThrIleAspArgAlaSerGlnValLeuGlyTyr 42
Db 56651 TGGAGGGCGTGGCGCGCTTCGCGAGACCTTCGACCGGTGCTCACGCTTTCGACCGG 56710
Qy 43 AspLeuArgTyrLeuIle-----AspThrGluGluAsp 53
Db 56711 GAGCTCCATCAGCCCTCTCGAGGTGATGGCGGCGAGCGCGGCGAGCAGCGTCTGTCG 56770
Qy 54 LysLeuAsnGlnThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyr 73
Db 56771 TTGCTGGACACAGACCGCTTCACCCAGCCGCGCTCTTTGCGTGGAGTACGCGCTGCGC 56830
Qy 74 ArgLeuLeuGlnGlnLysGlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGlu 93
Db 56831 GCGCTCTCCGTCGTCGGCGGTGGAGCGGAGCTGCTGCTGGCCATAGCCTCGCGCAG 56890
Qy 94 TyrSerAlaLeuValAlaSerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAla 113
Db 56891 CTGGTGGCGCGCTGCGTGGCGGTGTTCCTCTCGAGGACCGCTGCGCTTGTGTGTC 56950
Qy 114 LysArgGlyAlaTyrMetGluGlnAlaAlaProAlaAspSerGlyLysMetValAlaVal 133
Db 56951 GCGCGCGCGCTTGTATG---CAGCGCTGTCGCGCC---GGCGCGCGATGATATCGATC 57004
Qy 134 LeuAsnThrProValGluValIleGluAlaCysGlnLysAlaSerGluLeuGlyVal 153
Db 57005 GCGCGCGCGAGCGCGCGCTG-----GCTCGCGGTGGCGCGCGCGCGAGCGGTG 57055
Qy 154 ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleAlaGlyGluValValAla 173
Db 57056 GTGTGATCGCGCGAGTCAATGGCGCGAGCAGGTGTGTGATCGCGGCGCGGAGAAATTC 57115
Qy 174 ValAspArgAlaValGluLeuGlnGlnAlaGlyAlaLysArgLeuIleProLeuLys 193
Db 57116 GTGACGACGATCGCGCGCGTTCGCGCGCGCGGGGCG---CGAACCAACCGCTGCAT 57172
Qy 194 ValSerGlyProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThr 213
Db 57173 GTCTCGACGCGTTCCTCTCGCGCTCATGATCGATGCTGGAGCGCTTCGCGCGGTG 57232
Qy 214 LeuAlaGlnValSerPheSerAspPheThrCysProLeuValGlyAsnThrGluAlaAla 233
Db 57233 ACTGAGTCGGTACGTCACCGCGCGCTTCGATCGCGCTGGTGGAGCAACTGAGCGGGAAG 57292
Qy 234 ValMetGlnLysGluAspIleAlaGln---LeuLeuThrArgGlnValLysGluProVal 252
Db 57293 CCTTCACCGATAGGTGAGCGCGCGGTACTGCGGTGCTCACCGCGAGAGCGGTG 57352
Qy 253 ArgPheTyrGluSerIleGlyValMetGlnGlnAlaGlyIleSerAsnPheIleGluIle 272
Db 57353 CGCTTCGCGGACGAGTGAAGCGCTGCACGCGCGCGTGCAGCGCGCTTCTTCGTCGAGGTG 57412
Qy 273 GlyProGlyLysValLeuSerGlyPheVal 282
Db 57413 GGGCCGAAGCCGACGCTGCTCGCGCTTGTG 57442

RESULT 11

US-09-568-102-1

; Sequence 1, Application US/09568102

; Patent No. 6346404

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

Db 56891 CTGGTGGCCGCTCGTGGGGGTGTTCTCCCTCGAGACCGCGTGGCGCTTGGTGGTC 56950
 Qy 114 LysArgGlyAlaTyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaVal 133
 Db 56951 GCGCGCGCGCGTGGTATG---CAGCGCTGCCGCC---GGCGCGCGATGGTATCGATC 57004
 Qy 134 LeuAsnThrProValGluValIleGluGluAlaCysGlnLysAlaSerGluLeuGlyVal 153
 Db 57005 GCGCGCGCGAGCGCGACGTG-----GCTGCGCGGTGGCGCGACCGAGCGTTG 57055
 Qy 154 ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleAlaGlyGluValValAla 173
 Db 57056 GTGTGCGATCGCGCGAGTCAATGCGCGAGCAGGTGTGATCGCGCGCGCGAGAAATTC 57115
 Qy 174 ValAspArgAlaValGluLeuLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLys 193
 Db 57116 GTGCAGCAGATCGCGCGCGTTCGCGCGCGCGCGCGCG---CGAACCAACCGCTGCAT 57172
 Qy 194 ValSerGlyProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThr 213
 Db 57173 GTCTCGCACCGTTCACCTCGCGCTCATGTATCGATCGTGGAGCGTTCGCGCGGGTG 57232
 Qy 214 LeuAlaGlnValSerPheSerAspPheThrCysProLeuValGlyAsnThrGluAlaAla 233
 Db 57233 ACTGAGTCGTGACGTACCGCGCGCGCTTCGATCGCGTGTGAGCAACCTGAGCGGGAAG 57292
 Qy 234 ValMetGlnLysGluAspIleAlaGln---LeuLeuThrArgGlnValLysGluProVal 252
 Db 57293 CCCTGCACCATGAGTGACCGCGCGGGGTACTGGTGTGCTACCGCGAGAGCGGGTG 57352
 Qy 253 ArgPheTyrGluSerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIle 272
 Db 57353 CGTTTCGCGACGAGTGAAGCGCTGCACGCGCGGTGCGCGCGCTCTTCTGTCGAGGTG 57412
 Qy 273 GlyProGlyLysValLeuSerGlyPheVal 282
 Db 57413 GGGCCGAAGCGGCGCTGCTCGGCCTTGTG 57442
 RESULT 13
 US-09-568-480-1
 ; Sequence 1, Application US/09568480
 ; Patent No. 6355458
 ; GENERAL INFORMATION:
 ; APPLICANT: Schupp, Thomas
 ; APPLICANT: Ligon, James
 ; APPLICANT: Molnar, Istvan
 ; APPLICANT: Zirkle, Ross
 ; APPLICANT: Cvr, Devon
 ; APPLICANT: Goerlach, Joern
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 ; FILE REFERENCE: 4-30582A
 ; CURRENT APPLICATION NUMBER: US/09/568,480
 ; PRIOR FILING DATE: 2000-05-10
 ; PRIOR APPLICATION NUMBER: 09/335,409
 ; PRIOR FILING DATE: 1999-06-17
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 68750
 ; TYPE: DNA
 ; ORGANISM: Sorangium cellulosum
 US-09-568-480-1

Alignment Scores:
 Pred. No.: 1,13e-31 Length: 68750
 Score: 374.00 Matches: 101
 Percent Similarity: 47.93% Conservative: 38
 Best Local Similarity: 34.83% Mismatches: 135
 Query Match: 24.64% Indels: 16
 DB: 4 Gaps: 6

US-09-308-397-2 (1-306) x US-09-568-480-1 (1-68750)

Qy 3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22
 Db 56591 AAGTCGCTTCCTTCCTTCGCGCGCGAGCGCGAGTTCGCGCGATGGCGCGTGGTGG 56650
 Qy 23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
 Db 56651 TGGAGAGCGTGGCGCGGTTCGCGGAGACCTTCGACCGGTGCGTCACGCTTCGACCGG 56710
 Qy 43 AspLeuArgTyrLeuIle-----AspThrGluGluAsp 53
 Db 56711 GAGTCCATCAGCGCGCTCTCGAGGTGATGTGGGCCGAGCGCGGCGAGCAGCGTCGTCG 56770
 Qy 54 LysLeuAsnGlnThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyr 73
 Db 56771 TTGCTGGACCGAGCGCGGTTCACCCAGCGCGGTCTTTTCGCTGGAGTAGCGCTCGGCC 56830
 Qy 74 ArgLeuLeuGlnGlyLysGlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGlu 93
 Db 56831 GCGCTCTTCGCGTGGCGGTGGAGCGGAGTCTGCTGCGCATAGCTCGCGGAG 56890
 Qy 94 TyrSerAlaLeuValAlaSerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAla 113
 Db 56891 CTGTGCGCGCTCGTGGCGGGGTGTTCCTCCTCGAGGACCGCGTGGCTGGTGGTC 56950
 Qy 114 LysArgGlyAlaTyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaVal 133
 Db 56951 GCGCGCGCGCGTGTATG---CAGCGCTGCCGCC---GGCGCGCGATGGTATCGATC 57004
 Qy 134 LeuAsnThrProValGluValIleGluGluAlaCysGlnLysAlaSerGluLeuGlyVal 153
 Db 57005 GCGCGCGCGAGCGCGACGTG-----GCTGCGCGGTGGCGCGCGCGCGAGCGGTG 57055
 Qy 154 ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleAlaGlyGluValAlaAla 173
 Db 57056 GTGTGATCGCGCGAGTCAATGCGCGAGAGGTGTGATCGCGCGCGCGAGAAATTC 57115
 Qy 174 ValAspArgAlaValGluLeuLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLys 193
 Db 57116 GTGCAGCAGATCGCGCGCGGTTCGCGCGCGCGGGCGG---CGAACCAACCGCTGCAT 57172
 Qy 194 ValSerGlyProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThr 213
 Db 57173 GTCTCGCACCGTTCACCTCGCGCTCATGTATCGATCGTGGAGCGTTCGCGCGGGTG 57232
 Qy 214 LeuAlaGlnValSerPheSerAspPheThrCysProLeuValGlyAsnThrGluAlaAla 233
 Db 57233 ACTGAGTCGTGACGTACCGCGCGCGCTTCGATCGCGTGTGAGCAACCTGAGCGGGAAG 57292
 Qy 234 ValMetGlnLysGluAspIleAlaGln---LeuLeuThrArgGlnValLysGluProVal 252
 Db 57293 CCCTGCACCATGAGTGACCGCGCGGGGTACTGGTGTGCTACCGCGAGAGCGGGTG 57352
 Qy 253 ArgPheTyrGluSerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIle 272
 Db 57353 CGTTTCGCGACGAGTGAAGCGCTGCACGCGCGGTGCGCGCGCTCTTCTGTCGAGGTG 57412
 Qy 273 GlyProGlyLysValLeuSerGlyPheVal 282
 Db 57413 GGGCCGAAGCGGCGCTGCTCGGCCTTGTG 57442
 RESULT 14
 US-09-568-486-1
 ; Sequence 1, Application US/09568486
 ; Patent No. 6355459
 ; GENERAL INFORMATION:
 ; APPLICANT: Schupp, Thomas
 ; APPLICANT: Ligon, James
 ; APPLICANT: Molnar, Istvan
 ; APPLICANT: Zirkle, Ross
 ; APPLICANT: Cvr, Devon
 ; APPLICANT: Goerlach, Joern
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 ; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Alignment Scores:

Pred. No.: 1,138-31 Length: 68750
Score: 374.00 Matches: 101
Percent Similarity: 47.93% Conservative: 38
Best Local Similarity: 34.83% Mismatches: 135
Query Match: 24.64% Indels: 16
DB: 4 Gaps: 6

US-09-308-397-2 (1-306) x US-09-568-486-1 (1-68750)

QY 3 LysThrAlaPheLeuPheAlaGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22
Db 56591 AAGCTCGCCTTCTGTTCCCGGCGAGCGCGCGAGTCCCGCGCATGGCGGTG 56650
QY 23 TyrAspGlnTyrProIleValIysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
Db 56651 TGGGAGCGTGGCGCGGTTCGCGAGACCTTCGACCGGTGCTCAAGCTCTTCGACCGG 56710
QY 43 AspLeuArgTyrLeuLeu-----AspThrGluGluAsp 53
Db 56711 GAGCTCCATCAGCCGCTCTCGAGGTGATGGCGCGAGCCGCGAGCAGCGTCTCG 56770
QY 54 LysLeuAsnGlnThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyr 73
Db 56771 TTGCTGGACACCGCGTTCACCGCGCGGCTTTGGCTGGAGTACGCGCTGGCC 56830
QY 74 ArgLeuLeuGlnGluGlyGlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGlu 93
Db 56831 GCGCTCTTCGCTGCGGCGGTGAGCGCGAGCTCGCTGGCCATAGCTCGCGGAG 56890
QY 94 TyrSerAlaLeuValAlaSerGlnAlaLeuAspPheGluAspAlaValAlaValAla 113
Db 56891 CTGGTGGCGCGTGGCGGTGTGTCTCCCTCGAGGACCGCGTGGTGGTGTG 56950
QY 114 LysArgGlyAlaTyrMetGluGluAlaProAlaAspSerGlyLysMetValAlaVal 133
Db 56951 GCGCGCGCGCGGTGATG---CAGCGCTGCGCGCC---GCGCGCGCATGATCGATC 57004
QY 134 LeuAsnThrProValGluValIleGluGluAlaCysGlnLysAlaSerGluLeuGlyVal 153
Db 57005 GCGCGCGCGAGCGCGAGTGTG-----GCTCGCGCGGTGGCGCGCGACGCGTGTG 57055
QY 154 ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleAlaGlyGluValAlaVal 173
Db 57056 GTGTCGATCGCGCGAGTCAATCGGCGCGAGCGGTGTGATCGGCGCGCGCGAATTC 57115
QY 174 ValAspArgAlaValGluLeuLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLys 193
Db 57116 GTGACGACATCGCGCGCGGTTCGCGCGCGCGCGCGGCGGCG---CGAACCAACCGCTGCAT 57172
QY 194 ValSerGlyProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThr 213
Db 57173 GTCTCGACGGTTCATCTCGCGCTCATGATCGATGCTGGAGGCGTTCGCGCGGTG 57232
QY 214 LeuAlaGlnValSerPheSerAspPheThrCysProLeuValGlyAsnThrGluAlaVal 233
Db 57233 ACTGAGTCGGTGACGTACCGCGCGGTTCGATCGCGCTGCTGAGCAACCTGAGCGGGAAG 57292
QY 234 ValMetGlnLysGluAspIleAlaGln---LeuLeuThrArgGlnValLysGluProVal 252
Db 57292 CCCTGCACCGATAGGTGAGCGCGCGCGCGGTACTGGGTGCTGACGCGCGAGGCGGTG 57352

QY 253 ArgPheTyrGluSerIleGlyValMetGlnGluAlaGlyLysSerAsnPheIleGluIle 272
Db 57353 CGCTTCGCGAGCGAGTGAAGCGCTGACGCGCGCGGTGCGCGGCTCTTCGTCGAGGTG 57412
QY 273 GlyProGlyLysValLeuSerGlyPheVal 282
Db 57413 GGGCCGAGCGGAGCGCTGCTCGGCGCTGTG 57442

RESULT 15

US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkie, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Alignment Scores:
Pred. No.: 1,138-31 Length: 68750
Score: 374.00 Matches: 101
Percent Similarity: 47.93% Conservative: 38
Best Local Similarity: 34.83% Mismatches: 135
Query Match: 24.64% Indels: 16
DB: 4 Gaps: 6

US-09-308-397-2 (1-306) x US-09-568-472-1 (1-68750)

QY 3 LysThrAlaPheLeuPheAlaGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22
Db 56591 AAGCTCGCCTTCTGTTCCCGGCGAGCGCGCGAGTCCCGCGCATGGCGGTG 56650
QY 23 TyrAspGlnTyrProIleValIysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
Db 56651 TGGGAGCGTGGCGCGGTTCGCGAGACCTTCGACCGGTGCTCAAGCTCTTCGACCGG 56710
QY 43 AspLeuArgTyrLeuLeu-----AspThrGluGluAsp 53
Db 56711 GAGCTCCATCAGCCGCTCTCGAGGTGATGGCGCGAGCCGCGAGCAGCGTCTCG 56770
QY 54 LysLeuAsnGlnThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyr 73
Db 56771 TTGCTGGACACCGCGTTCACCGCGCGGCTTTGGCTGGAGTACGCGCTGGCC 56830
QY 74 ArgLeuLeuGlnGluGlyGlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGlu 93
Db 56831 GCGCTCTTCGCTGCGGCGGTTCGCGCGCGCGGTGAGTCCGCTGGAGTACGCGCTGGCC 56890
QY 94 TyrSerAlaLeuValAlaSerGlyAlaLeuAspPheGluAspAlaValAlaValAla 113
Db 56891 CTGGTGGCGCGTGGCGGTGTGTCTCCCTCGAGGACCGCGTGGTGGTGTG 56950
QY 114 LysArgGlyAlaTyrMetGluGluAlaProAlaAspSerGlyLysMetValAlaVal 133
Db 56951 GCGCGCGCGCGGTGATG---CAGCGCTGCGCGCC---GCGCGCGCATGATCGATC 57004
QY 134 LeuAsnThrProValGluValIleGluGluAlaCysGlnLysAlaSerGluLeuGlyVal 153

```
Db 57005 GCCGCGCGGAGCGGACGTTG-----GCTGCCCGGTGCGCGCGACGCGGTTG 57055
Qy 154 ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleAlaGlyGluValAla 173
Db 57056 GTGTGATCGCGGAGTCAATGGCGGAGCGAGTGTGATCGCGGCGCGGAGAAATTC 57115
Qy 174 ValAspArgAlaValGluLeuGlnGluAlaGlyAlaIysArgLeuIleProLeuLys 193
Db 57116 GTGACGAGATCGCGGCGGCTTCGCGCGCGCGGGGCGG---CGAACCAACCGCTGCAT 57172
Qy 194 ValSerGlyProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThr 213
Db 57173 GTCTCGACGCGTCCACTCGCGCTCATGATCCGATGCTGAGCGGTTCCGGCGGGTG 57232
Qy 214 LeuAlaGlnValSerPheSerAspPheThrCysProLeuValGlyAsnThrGluAla 233
Db 57233 ACTGAGTCGCTGACGTACCGCGCGGCGCTTCGATCGCGCTGTGAGCAACCTGAGCGGGAAG 57292
Qy 234 ValMetGlnLysGluAspIleAlaGln---LeuLeuThrArgGlnValLysGluProVal 252
Db 57293 CCCTGCACCGATGAGGTGAGCGCGCGGGTTACTGGGTGCGTCACCGCGGAGAGCGGTTG 57352
Qy 253 ArgPheTyrGluSerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIle 272
Db 57353 CGCTTCGCGACGAGGTGAAGGCGCTGCACGCGCGCGGTCGCGGCGCTCTTCGTCGAGGTG 57412
Qy 273 GlyProGlyLysValLeuSerGlyPheVal 282
Db 57413 GGGCGGAGCGGAGCGCTGCTCGGCGCTTGTG 57442
```

Search completed: June 11, 2003, 21:54:52
Job time : 1805 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 11, 2003, 21:03:06 ; Search time 155 seconds
(without alignments)
2762.958 Million cell updates/sec

Title: US-09-308-397-2

Perfect score: 1518

Sequence: 1 MTKTAFLPAGCAQVLMGR.....QTAHLAHEVDQASLVALLEK 306

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 870385 seqs, 699768693 residues

Total number of hits satisfying chosen parameters: 1740770

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USPTO.spoc1/US09308397/runat 06062003 112350 28979/app query.fasta_1.455
-DB=Published Applications NA -QMT=fastap -SUFFIX=rpnb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODS=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09308397 @CGN 1 1 104 @runat 06062003 112350 28979
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FCAPOP=6 -FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1514	99.7	921	10	US-09-815-242-9482 Sequence 9482, Ap
2	937	61.7	942	10	US-09-815-242-6823 Sequence 6823, Ap
3	920	60.6	3656	10	US-09-070-927A-5777 Sequence 577, App
4	643.5	42.4	930	10	US-09-815-242-6028 Sequence 6028, Ap

5	636	41.9	939	10	US-09-815-242-6878 Sequence 6878, Ap
6	636	41.9	1830121	9	US-10-329-960-1 Sequence 1, Appli
7	627.5	41.3	950	10	US-09-974-300-1143 Sequence 1143, Ap
8	627.5	41.3	960	10	US-09-974-300-1086 Sequence 1086, Ap
9	556.5	36.7	906	10	US-09-815-242-4405 Sequence 4405, Ap
10	556.5	36.7	936	10	US-09-815-242-8032 Sequence 8032, Ap
11	556.5	36.7	939	10	US-09-815-242-7796 Sequence 7796, Ap
12	550.5	36.3	6251	7	US-08-781-986A-25 Sequence 25, Appli
13	421	27.7	6021	10	US-09-070-927A-458 Sequence 458, App
14	416.5	27.4	930	10	US-09-815-242-7193 Sequence 7193, Ap
15	416	27.4	532	10	US-09-974-300-5552 Sequence 5552, Ap
16	414.5	27.3	4209	9	US-09-712-363-20 Sequence 20, Appli
17	379	25.0	640681	10	US-09-790-988-1 Sequence 1, Appli
18	376.5	24.8	50937	9	US-09-808-880-1 Sequence 1, Appli
19	374	24.6	68750	9	US-10-014-717-1 Sequence 1, Appli
20	372.5	24.5	15872	9	US-09-860-846-1 Sequence 1, Appli
21	372.5	24.5	15872	9	US-09-888-384B-1 Sequence 1, Appli
22	372.5	24.5	15872	9	US-09-836-821-1 Sequence 1, Appli
23	372.5	24.5	15872	10	US-09-861-289-1 Sequence 1, Appli
24	371	24.4	4851	9	US-09-712-363-116 Sequence 116, App
25	348.5	23.0	8268	9	US-10-331-061-83 Sequence 83, Appli
26	348.5	23.0	37895	9	US-10-331-061-1 Sequence 1, Appli
27	348	22.9	1380	9	US-10-074-045-24 Sequence 24, Appli
28	343.5	22.6	7959	9	US-10-331-061-77 Sequence 77, Appli
29	343.5	22.6	19227	9	US-10-331-061-13 Sequence 13, Appli
30	343.5	22.6	40138	9	US-10-331-061-12 Sequence 12, Appli
31	340.5	22.4	1278	9	US-10-124-800-9 Sequence 9, Appli
32	340.5	22.4	8730	9	US-10-124-800-1 Sequence 1, Appli
33	334	22.0	1030	9	US-09-735-056-2 Sequence 2, Appli
34	324	21.3	13842	9	US-09-860-846-30 Sequence 30, Appli
35	324	21.3	13842	9	US-09-888-384B-30 Sequence 30, Appli
36	324	21.3	13842	9	US-09-836-821-30 Sequence 30, Appli
37	324	21.3	13842	10	US-09-861-289-30 Sequence 30, Appli
38	324	21.3	36778	9	US-09-860-846-5 Sequence 5, Appli
39	324	21.3	36778	9	US-09-836-821-5 Sequence 5, Appli
40	324	21.3	36778	10	US-09-861-289-5 Sequence 5, Appli
41	324	21.3	37948	9	US-09-988-384B-5 Sequence 1, Appli
42	270	17.8	7692	9	US-09-735-056-1 Sequence 1, Appli
43	254	16.7	7692	9	US-09-836-705-45 Sequence 45, Appli
44	236	15.5	34203	9	US-09-836-705-1 Sequence 1, Appli
45	236	15.5	34203	9	US-09-836-705-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-815-242-9482
; Sequence 9482, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9482
; LENGTH: 921
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(921)
US-09-815-242-6822

Alignment Scores:
Pred. No.: 1,93e-175 Length: 921
Score: 1514.00 Matches: 305
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.67% Mismatches: 0
Query Match: 99.74% Indels: 0
DB: 10 Gaps: 0

US-09-308-397-2 (1-306) x US-09-815-242-9482 (1-921)

QY 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
Db 1 ATGACTAAACAGCCCTTTTATTTGCTGTCAGAGTGCCAGATCTAGGGATGGACGG 60
QY 21 AspPheTyrAspGlnTyrProLysValLysGluThrLysLeuAspArgAlaSerGlnValLeu 40
Db 61 GATTTCATGATCAGTATCCGATGTGCAAGAACCATGATCGAGCGAGTCAGGTGCTA 120
QY 41 GlyTyrAspLeuArgTyrLeuLysLeuAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60
Db 121 GGTATGATTTGCTTATCTCATCGATACGAAAGAGCAACAATCAATCAGACCCGCTAT 180
QY 61 ThrGlnProLysLeuAlaThrSerValAlaLysLeuLysLeuGlnGluGlyGly 80
Db 181 ACGCAACAGCCCATTCAGGATTCGGTTGCTATCTACCGTTTATTGCAAGAAAGGCG 240
QY 81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyLysSerAlaLeuValAlaSer 100
Db 241 TATCAGCCTGATATGGTTGCTGTTGCTCTTGGAGATACCTCTGCTTGGTGGCAAC 300
QY 101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
Db 301 GCGCCCTTGGATTGGAAGATGCGGTTGCTTGGTAGCTAAGCGTGGAGCCTATATGGA 360
QY 121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValAlaValLeuAsnThrProValGluVal 140
Db 361 GAAGCGGCTCTGCTGATCTGCAAGATGTTAGCAGTCTCTCATACGCCAGTAGAGTTC 420
QY 141 IleGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProLysThrAsn 160
Db 421 ATTGAAGACCGCTGTCAAAAGCTTCTGAACCTGGAGTGGTTACTCCAGCCCACTATAAC 480
QY 161 ThrProLysGlnLeuValLysLeuGlyValValAlaValAlaValAspArgAlaValGluLeu 180
Db 481 ACACCTGCACAAATCGTCATGCTGGAGAGTGGTTGCAAGTGTGATCGAGCGGTGAACCT 540
QY 181 LeuGlnGluAlaGlyAlaLysArgLeuLeuProLeuLysValSerGlyProPheHisThr 200
Db 541 TTGCAAGAACGAGGTGCAACCGCTTGAATTCCTCTTAAGGTGTGAGTCCCTTTCACACC 600
QY 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaValSerPheSer 220
Db 601 GCTCTCCTTGAACCTGTACGCCAAGAACTAGCTGAAACTCTAGCTCAGGTAAAGTTTTC 660
QY 221 AspPheThrCysProLeuValGlyAsnThrGluAlaValMetGlnLysGluAspPhe 240
Db 661 GATTTCATGTCCTCCCTAGTCGCAATACAGAAGCTGCTGTGATGCAAAAGAGACAT 720

QY 241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
Db 721 GCTCAGCTCTTGCAGCGCTCAGTCAAGAACCCCTTCTCTATGAAGATTTGGGCTC 780
QY 261 MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280
Db 781 ATGCAAGAGCAGGCATTAAGCACTTTATCGAGATTGGACCGGGGAAAGTTTGTCAAGT 840
QY 281 PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerIle 300
Db 841 TTTGTTAAAAAATTGATCGAACTGCTCACTAGCTCATGTGGAAGATCAAGCGAGTTTA 900
QY 301 ValAlaLeuLeuGluLys 306
Db 901 GTAGCACCTTTTAGAATAA 918

RESULT 2
US-09-815-242-6823
; Sequence 6823, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6823
; LENGTH: 942
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(942)
US-09-815-242-6823

Alignment Scores:
Pred. No.: 5,94e-105 Length: 942
Score: 937.00 Matches: 187
Percent Similarity: 75.17% Conservative: 40
Best Local Similarity: 61.92% Mismatches: 75
Query Match: 61.73% Indels: 0
DB: 10 Gaps: 0

US-09-308-397-2 (1-306) x US-09-815-242-6823 (1-942)

QY 3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyArgAspPhe 22
Db 16 AAAACAGCGATTTTATTATTGTTGGAAGAGCCCAAGTATCAAGGATGGGTGAAGATTA 75
```

QY 23 TyrAspGlnTyrProIleValIleGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
Db 76 TATCACCAGAACGATGTTCCGGAACTTTCCGATGAAGCAAGTCATATCTTAGGTTAT 135
QY 43 AspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyrThrGln 62
Db 136 GAGATGGCAGCACTTTGTTTACTGAAATGAACGCTTAAATGAACAGATATACGCAA 195
QY 63 ProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLysGlyTyrGln 82
Db 196 CTGCTATTATTAACAGTCAGTGTCCCATTTTACCGTCTTTTGCAACAAAAAGGACTAACG 255
QY 83 ProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSerGlyVala 102
Db 256 CTGATGTCGTAGCGGTTTAACTTAGGGAATATACGCTTTGGTTGCCAGCGGGGCT 315
QY 103 LeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluGluAla 122
Db 316 TTGCGCTTTTTCAGAACGACGTGGCTTGGTCCAAAGCGCGGTCAATCATGACAGAAACA 375
QY 123 AlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGlu 142
Db 376 GCACCAACAGNACTGGCAAAATGGTTGCTGTCATGAATGCTGAGCGTGAAGTAATTGAG 435
QY 143 GluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThrPro 162
Db 436 AAAGCGCTGCCAAGAACCGCAGTCTTTCGGAATTTGGCTCCAGCAAAATTAATACACCA 495
QY 163 AlaGlnIleValIleAlaGlyValValAlaValAlaAspArgAlaValGluLeuGln 182
Db 496 CAACAAATCGTGAATGGTGGTGGAGGTTGCTGCTGTGATCAAGCGATGACACTTCTCAA 555
QY 183 GluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeu 202
Db 556 GAAGCTGGTGTGAACGAATGATTCGTTAAATGTGAGTGGCCCTTTCATACGCGCTG 615
QY 203 LeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPhe 222
Db 616 TTACCAACAGCATCAAAAAATTTGGCTCAGGATTTAGCAAAATTTGAACCTTTCAAACGATG 675
QY 223 ThrCysProLeuValGlyAsnThrGluAlaValMetGlnLysGluAspIleAlaGln 242
Db 676 CAATTCCTGTCATTAGTAATACGACTGCCGAATTTATGCCCAAGAGGCAATCAAGCG 735
QY 243 LeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGln 262
Db 736 TTATTGGAAGAACAGTCTATCTCGGTACGTTTGAAGACAGTATCGAAACGATGAAG 795
QY 263 GluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGlyPheVal 282
Db 796 GCTATGAACGTGGGACGATGATGAGTTGGTCCAGGAAAAACATTAACTGGTTTGT 855
QY 283 LysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeuVala 302
Db 856 AAAAAATTGACAAAAACAATTGAATGCCCGTGTGGAAGATGTTGCAACATTAAACAGAA 915
QY 303 LeuLeu 304
Db 916 ACGTTA 921

RESULT 3

US-09-070-927A-577
; Sequence 577, Application US/09070927A
; Patent No US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue

CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICANT NUMBER: US/09/070, 927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 577:
SEQUENCE CHARACTERISTICS:
LENGTH: 3656 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 577:

US-09-070-927A-577

Alignment Scores:
Pred. No.: 5,21e-102 Length: 3656
Score: 920.00 Matches: 186
Percent Similarity: 75.25% Conservative: 39
Best Local Similarity: 62.21% Mismatches: 73
Query Match: 60.61% Indels: 1
DB: 10 Gaps: 0

US-09-308-397-2 (1-306) x US-09-070-927A-577 (1-3656)

QY 3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22
Db 2753 AAAACAGCGATTTTATTAGTGACACAGGAGCCAGTATCAAGGGATGGTGAAGAATTA 2812
QY 23 TyrAspGlnTyrProIleValIleGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
Db 2813 TATCACCAGAACGATGTTCCGGAACTTTTCGATGAAGCAAGTCATATCTTAGGTTAT 2872
QY 43 AspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyrThrGln 62
Db 2873 GAGATGGCAGCACTTTGTTTACTGAAATGAACGTTAAATGAACAGATATACGCAA 2932
QY 63 ProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLysGlyTyrGln 82
Db 2933 CTGCTATTATTAACAGTCAGTGTCCCATTTTACCGTCTTTTTCACAAAAAGGACTAACG 2992
QY 83 ProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSerGlyVala 102
Db 2993 CTGATGTCGTAGCGGTTTAACTTAGGGAATATACGCTTTGGTTGCCAGCGGGGCT 3052
QY 103 LeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluGluAla 122
Db 3053 TTGCGCTTTTTCAGAACGACGTGGCTTGGTCCAAAGCGCGGTCAATCATGACAGAAACA 3112
QY 123 AlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGlu 142
Db 3113 GCACCAACAGGAACTGGCAAAATGGTTGCTGTTATGAATGCTGAGCGTGAAGTAATTGAG 3172

```
QY 143 GluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThrPro 162
Db 3173 AAAGCCTGCCAAGAAGCAGCGCTTCGGAATTCGGCTCCAGCAATATAACACCA 3232
QY 163 AlaGlnLeuValIleAlaGlyGluValAlaValAlaAspArgAlaValGluLeuGln 182
Db 3233 CAACAATCGATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3292
QY 183 GluAlaGlyAlaLysArgLeuLeuProLeuLysValSerGlyProPheHisThrAlaLeu 202
Db 3293 GAAGCTGGTGTGAAGCGAATGATTCCTGTAATGAGTGGTGGTGGTGGTGGTGGT 3352
QY 203 LeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPhe 222
Db 3353 TTACACCGAGCTCAAAAATTTGGCTCAGATTTAGCAAAATGAACTTCAACGATG 3412
QY 223 ThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIleAlaGln 242
Db 3413 CAATTCCTGTCTAGTAATACGACTGCCGAATTTATGCCCAAGAGGCAATTCAGCG 3472
QY 243 LeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGln 262
Db 3473 TTATTTGAAAGAGCAAGTCATGCTGCGGTACGTTTTCGAAGACAGATTCGAAACGATGAAG 3532
QY 263 GluAlaGlyIleSerAsnPheIleGluIle-GlyProGlyLysValLeuSerGlyPheVa 282
Db 3533 GCTATGAACGTAGGAACGATGATTTGAAGTTGGTCCAGGGRAACATTAACYGTTTGT 3592
QY 282 LysIleValSerGlnThrAlaHisLeuAlaHisValGluAlaSerLeu 300
Db 3593 TAAAAAATTACCAAAACAAATTGAATGCACCGTGTGGAAGATGTTGCCACATTA 3647
```

RESULT 4

```
US-09-815-242-6028
; Sequence 6028, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6028
; LENGTH: 930
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
```

LOCATION: (1)....(930)
US-09-815-242-6028

Alignment Scores:
Pred. No.: 4, 1e-69 Length: 930
Score: 643.50 Matches: 141
Percent Similarity: 64.84% Conservative: 60
Best Local Similarity: 45.48% Mismatches: 102
Query Match: 42.39% Indels: 7
DB: 10 Gaps: 4

US-09-308-397-2 (1-306) x US-09-815-242-6028 (1-930)

QY 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
Db 1 ATGACGCAATTTGCAATTGTTCTCCAGCAGGGTTCCTCAACCGTTGGAAATGCTGGCT 60
QY 21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
Db 61 GATATGGCGCGGAGCTATCCAATTTGTCGAAGAACGTTTGTGGAAGCTTCTGCGCGCGCTG 120
QY 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGlu-----AspLysLeuAsnGln 57
Db 121 GGCTACGACCTGTGGGCGCTG-----ACCCAGCAGGGGCCAGCTGAAGAACTGAATAAAA 174
QY 58 ThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGln 77
Db 175 ACCTGGCAAACTCAGCCTGCGCTGTGACTGCATCTGTTGCGTGTATCGCGTATGGCAG 234
QY 78 GluLysGlyTyrGln---ProAspMetValAlaGlyLeuSerLeuGlyGlyTyrSerAla 96
Db 235 CAGCAGGCGGTAAAGCACCAGCAATGATGCCGCTCACAGCCTGGGGGAATATCTCCGCG 294
QY 97 LeuValAlaSerGlyValAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGly 116
Db 295 CTGGTTTGGCTGGTGTGATTGCTGCTGATGCGGTGCGCTGCTGTTGAGATCGCGGC 354
QY 117 AlaTyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThr 136
Db 355 AAGTTTCATGCAAGAGCCGCTACCGGAAGCAGCGCGCTATGGCGGCAATCATCGGTCTG 414
QY 137 ProValGluValIleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrPro 156
Db 415 GATGATGCGTCTATTGCGAAGCGTGTGAAGAGCTGCAGAAGGTGAGTCTGTTCTCCG 474
QY 157 AlaAsnTyrAsnThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArg 176
Db 475 GTAAACTTTAACTCTCCGGGACAGTGCTTATTGCGGTCATATAAGAGCGGTTGAGCGT 534
QY 177 AlaValGluLeuLeuGlnGluAlaGlyValAlaLysArgLeuIleProLeuLysValSerGly 196
Db 535 GCTGCGCTGCTGTAAAGCGCGGCGCAAAACGCGGCTGCGCTTACCAAGTACGCGTA 594
QY 197 PropPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGln 216
Db 595 CCGTCTCACTGTGCGCTGATGAACACGACGACCAAACTGGCAGTAGAATAGCGGAAA 654
QY 217 ValSerPheSerAspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGln 236
Db 655 ATCACCTTTAAACGACCAACAGTTCCTGTTGTGAATAACGTTGATGTGAATGCCAAACC 714
QY 237 LysGluAsp---IleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyr 255
Db 715 AATGCTGATGCCATCCGTCAGCAGCTGCTGATGATTAACCCGGTTCAGTGGACG 774
QY 256 GluSerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGly 275
Db 775 AAGTCTGTTGATACATGGCAGCGCAAGCGGTAGAACATCTCTATGAAGTCCGCCCGGCG 834
QY 276 LysValLeuSerGlyPheValLysIleAspGlnThrAlaHisLeuAlaHisValGlu 295
Db 835 AAAGTGCCTTACTGGCTGACGAACGCAATTGTCGACACCCCTGACCGCTCGCGCTGAAC 894

QY 296 AspGlnAlaSerLeuValAlaLeuLeuGlu 305
 Db 895 GAACCTTCAGCGATGCGCGCGCTCGAG 924

RESULT 5

US-09-815-242-6878

; Sequence 6878, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlssen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6878

; LENGTH: 939

; TYPE: DNA

; ORGANISM: Haemophilus influenzae

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)....(939)

US-09-815-242-6878

Alignment Scores:

Pred. No.:	3,438-68	Length:	939
Score:	636.00	Matches:	142
Percent Similarity:	63.43%	Conservative:	54
Best Local Similarity:	45.95%	Mismatches:	109
Query Match:	41.90%	Indels:	4
Db:	10	Gaps:	3

US-09-308-397-2 (1-306) X US-09-815-242-6878 (1-939)

QY 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
 Db 1 ATGAAAAAATTCGCAATGGTCTCCAGGTCAAGCTCCCAACTGCCGATCGTCTGCT 60
 QY 21 AppPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
 Db 61 GATCTTGCACATGAATATCCATCGTTATGAAACATTTAAACCAAGCATCTGTATCGCGCTT 120
 QY 41 GlyTyrAspLeuArgTyrIleLeuAspThr---GluGluAspLysLeuAsnGlnThrArg 59
 Db 121 GGTGTAAGATTATGTATCTTCTTCAACAGGTCCAGCTGAAGAACTTAAATAAACCTGG 180
 QY 60 TyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLys 79
 Db 181 CAAACTCAGCCGCACTTTAGCTGCTTCAGTGGCTATTTATCGGTATGGAAGAAAAA 240

QY 80 GlyTyrGln-----ProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeu 97
 Db 241 TTTCTCTCAATTAACACCAAGATGAGGAGGTATAGCTTAGGTAGGAGTATCTTGGGTTA 300
 QY 98 ValAlaSerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAla 117
 Db 301 GTTTGTGCTGGGTGTTGGATTTCGAAGATGCGATTAAATATTAGTGAATTGCGCGAAAA 360
 QY 118 TyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrPro 137
 Db 361 TTAATGCAACAAAGCTGTGCTGAAGCACTGGCGCAATGTATGCAATCAATCGTTTAGAT 420
 QY 138 ValGluValIleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAla 157
 Db 421 AATGAAGCAATATTAAATCTTTCGAAACCAAGCAGAGGAGGCGAAGTGTATCTGCGGTG 480
 QY 158 AsnTyrAsnThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAla 177
 Db 481 AACTTTAACTACCGGGTCAAGTAGTTATTGCGGGTGCAGAAAGCTGCAGTTGAGCGTGG 540
 QY 178 ValGluLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyPro 197
 Db 541 GCTGCATTATGTAAGAAGCAGGCGGAAACCTGCATTCCTTACCTGTAGCGTACCT 600
 QY 198 PheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnVal 217
 Db 601 TCTCACTGTGCATTAAATGAACCTGCAGCGGAGCAATTAGCGGTAAACATTGAGAAATAT 660
 QY 218 SerPheSerAspPheThrCysProLeuValGlyAsnThrGluAla---AlaValMetGln 236
 Db 661 CAATTAATACACCAACATATCGGTATTAAATAACGTTGATGTGAAGCTGAAACTGAA 720
 QY 237 LysGluAspIleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGlu 256
 Db 721 GGCACCGAAATTCGTACCGCACTTGTGCGTCAGTTATATAGTCCAGTTCGTTGGACTGAA 780
 QY 257 SerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLys 276
 Db 781 ACAGTTGAAAAAATGCGCAAGATGCGGTTCTAGTCTTGTCTGAAGTGGCGCCAGTAAA 840
 QY 277 ValLeuSerGlyPheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAsp 296
 Db 841 GTATTAAATGTTTAAACCAACGCAATGTTGGGTGATTTACAGCAATATCTCTAAATGAT 900
 QY 297 GlnAlaSerLeuValAlaLeuLeuGlu 305
 Db 901 GTTGCATCATTCATGCGGTAGAGAA 927

RESULT 6

US-10-329-960-1/C

; Sequence 1, Application US/10329960

; Publication No. US20030099277A1

; GENERAL INFORMATION:

; APPLICANT: Fleischmann et al.

; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fra

; FILE REFERENCE: PB186P1

; CURRENT APPLICATION NUMBER: US/10/329,960

; PRIOR FILING DATE: 2003-01-02

; PRIOR APPLICATION NUMBER: US 09/643,990

; PRIOR FILING DATE: 2000-08-23

; PRIOR APPLICATION NUMBER: US 08/487,429

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: US 08/426,787

; PRIOR FILING DATE: 1995-04-21

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 1830121

; TYPE: DNA

; ORGANISM: Haemophilus influenzae

; FEATURE:

; NAME/KEY: misc_feature

LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c

Db 841 TTGTGAGGCTGTGAAGAAAGTGAACCGCAGAGTCAAAACAGTGTGTATCAGAC 897

RESULT 9

US-09-815-242-4405

; Sequence 4405, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4405

; LENGTH: 906

; TYPE: DNA

; ORGANISM: Staphylococcus aureus

US-09-815-242-4405

Alignment Scores:

Pred. No.:	1.67e-58	Length:	906
Score:	556.50	Matches:	122
Percent Similarity:	59.06%	Conservative:	54
Best Local Similarity:	40.94%	Mismatches:	117
Query Match:	36.66%	Indels:	5
DB:	10	Gaps:	4

US-09-308-397-2 (1-306) x US-09-815-242-4405 (1-906)

Qy	1	MethrlysthrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg	20
Db	1	ATGAGTAAACAGCAATATTTTCGGGCAAGAGTGCACCAAAAGTTGGTATGGCACAA	60
Qy	21	AspPheTyrAspGlnTyrProileVallysluThrleaspArgAlaSerGlnValleu	40
Db	61	GATTTATATACACACAGCATCAAGCAACTGAAATTTTAACCTCAGCAGCAACACAGTTA	120
Qy	41	GlyTyrAspLeuAtgTyrLeuileAspThrGluGluAsp---LysLeuAsnGlnThrArg	59
Db	121	GACTTTGATATTTTAGACACAATGTTTACTCAGCAAGCGCAATTTGGTGAAACTGAA	180
Qy	60	TyrThrGlnProAlaileLeuAlaThrSerVallelalleTyrArgLeuLeuGlnGluLys	79
Db	181	AACAGCAACAGCTTTATTGACGCATAGCTCGCGCTTATTAGCGCGCTA-----AAA	234
Qy	80	GlyTyrGlnProAspMetValalaGlyLeuSerLeuGlyGluTyrSerAlaLeuValala	99
Db	235	AATTGATCCGATTTTACTATGGGCATAGTTTAGTGAAATTTCAAGTTTAGTTGCA	294
Qy	100	SerGlyAlaLeuAspPheGluAspAlaValalaLeuValalaLysArgGlyAlaTyrMet	119

Db	295	GCTGAGCTATTATCATTTGAAGATCAGTTAAATTTGTAGAAAACGCTGGTCAATTAATG	354
Qy	120	GluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGlu	139
Db	355	GGCAGCAATTTCTTACTGCTAGGAACATGGCTGCAGTAGTTGGGCTTAGATTTTGAT	414
Qy	140	ValleleGluGluAlaCysGlnLys---AlaSerGluLeuGlyValValThrProAlaAsn	158
Db	415	AAAGTCGATGAATTTGTAAAGTCATTATCATCTGATGACAAATATTTGAAACGACAAAC	474
Qy	159	TyrAsnThrProAlaGlnleValleAlaGlyGluValValAlaValAlaAspArgAlaVal	178
Db	475	ATTAATTTGCCAGGTCAAATTTGTTTCAGGTCACAAAGCTTTAATTCATGCTAGTA	534
Qy	179	GluLeuLeuGlnGluAlaGlyAlaLysArgLeuileProLeuLysValSerGlyProPhe	198
Db	535	GAAAAGGTAAATCATTTAGGTGCAAAACGTCATCGCTTTAGCAGTAGTCTGCGCCATTC	594
Qy	199	HisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSer	218
Db	595	CATTATCGCTAATGAAAGTGATTGAAGAGATTTTCAAGTTATATTAATCAATTTGAA	654
Qy	219	PheSerAspPheThrCysProLeuValGlyAsnThrGluAla---AlaValMetGlnLys	237
Db	655	TGGCATGATGCTAAGTTTCTCTAGTTCAAAATGTAATGCGCAAGGTGAAACTGACAAA	714
Qy	238	GluAspIleAlaGlnLeuThrArgGlnVallysluThrProValArgPheTyrGluSer	257
Db	715	GAGTAATTAATCTAATATGTCACAAATATATTATTCACAGTACAAATTTAATCACTCA	774
Qy	258	IleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluileGlyProGlyLysVal	277
Db	775	ACAGAATGCTAATAGACCAAGGTGTGATCATCTTTTATTGAATTTGGTCTCGAAAAGTT	834
Qy	278	LeuSerGlyPheVallysluLysIleAspGlnThrAlaHisLeuAlaHisValGlu	295
Db	835	TTGTCTGGCTTAATTAATAAAATAAATAGAGATGTAAGTTAATCAATCAATTCAA	888

RESULT 10

US-09-815-242-8032

; Sequence 8032, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

```
; SEQ ID NO 8032
; LENGTH: 936
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(936)
US-09-815-242-8032

Alignment Scores:
Pred. No.: 1.75e-58 Length: 936
Score: 556.50 Matches: 122
Percent Similarity: 59.06% Conservativeness: 54
Best Local Similarity: 40.94% Mismatches: 117
Query Match: 36.66% Indels: 5
DB: 10 Gaps: 4

US-09-308-397-2 (1-306) x US-09-815-242-8032 (1-936)
QY 1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
DB 10 ATGAGTAAACAGCAATTTATTTCCGGACAGGTGCCCAAAAGTTGGTGGCACA 69
QY 21 AspPheTyrAspGlnTyrProileValLysGluThrIleAspArgAlaSerGlnValLeu 40
DB 70 GATTATATAACAACAACGATCAACGAACTGAAATTTTAACTTCAGCAGCAACACGTTA 129
QY 41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAsp---LysLeuAsnGlnThrArg 59
DB 130 GACITTGATATTTAGACACATGTTTACTGACGACGCGCAATTTGGTGAACCTGAA 189
QY 60 TyrThrGlnProAlaIleLeuAlaThrSerValIleLeuTyrArgLeuLeuGlnGluLys 79
DB 190 AACACGCAACAGCTTTATTGACGCATAGCTCGCGCTTATTAGCGCGCTA-----AAA 243
QY 80 GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAla 99
DB 244 AATTGATCCGATTTTACTATCGGCATAGTTAGTGAATATTCAGTTTAGTTGCA 303
QY 100 SerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMet 119
DB 304 GCTGACGTATTATCATTTGAAGTGCAGTTAAATTTGTAAGAAACGTCGCTCAATTAATG 363
QY 120 GluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGlu 139
DB 364 CGCAAGCATTTCTACTGTGTAGGAGCATGCTGCTGATTTGGCTTAGATTTTGAT 423
QY 140 ValIleGluGluAlaCysGlnLys---AlaSerGluLeuGlyValValThrProAlaAsn 158
DB 424 AAAGTCGATGAATTTGTAAGTCATTCATCTGATGACAAATAATTGAACAGCAAAAC 483
QY 159 TyrAsnThrProAlaGlnIleValIleAlaGlyGluValValAlaValAlaAspArgAlaVal 178
DB 484 ATTAATTCGCCAGGTCAATTTGTTTTCAGGTACAAAGCTTAATTTAGTACGTAGTA 543
QY 179 GluLeuLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPhe 198
DB 544 GAAAAAGGTAAATCATTTAGGTGCAAAACGTCATGCTTTAGCAGTATCTGSCCATTC 603
QY 199 HisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSer 218
DB 604 CATTCATCGCTTAATGAAGAGTGAATGAAGAGATTTTTCAGAGTTATATTAATCAATTGA 663
QY 219 PheSerAspPheThrCysProLeuValGlyAsnThrGluAla---AlaValMetGlnLys 237
DB 664 TGGCATATGCTAAGTTTCTGTAGTTCAAAATGTAATGCGCAAGGTGAACCTGACAAA 723
QY 238 GluAspIleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSer 257
DB 724 GAAGTAATTAATCAATCAATTTGGTCAACAATTTATTTACCAAGTACAAATTCATTAAC 783
QY 258 IleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluLeuGlyProGlyLysVal 277
DB 884 TGGCGCTGCTCCAGATGTCCTCAAGAGCGCTGAACAGACCCAGACCCAGCCG 195
QY 64 AlalleLeuAlaThrSerValAlaIleTyrArgLeu---LeuGlnGluLysGlyTyrGln 82
DB 844 TTGCTGCTGCTTAATTAATAAATAATAATAGATGTTAAGTTTAACATCAATCAAA 897

RESULT 11
US-09-815-242-7796
; Sequence 7796, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7796
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(939)
US-09-815-242-7796

Alignment Scores:
Pred. No.: 1.76e-58 Length: 939
Score: 556.50 Matches: 126
Percent Similarity: 62.54% Conservativeness: 51
Best Local Similarity: 44.52% Mismatches: 103
Query Match: 36.66% Indels: 3
DB: 10 Gaps: 3

US-09-308-397-2 (1-306) x US-09-815-242-7796 (1-939)
QY 5 AlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyrAsp 24
DB 16 GCATTCGCTTCTCCCTGCGCAGGTTCGCAATCCCTCGCATCTCGCGCGCTCGGCGGCC 75
QY 25 GlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyrAspLeu 44
DB 76 CAGCAGGCGCTGTGGCGGATACCTTCGCCAGGCTTCGCGCGCTCGGTACGACCTT 135
QY 45 ArgTyrLeuIleAspThr---GluGluAspLysLeuAsnGlnThrArgTyrThrGlnPro 63
DB 136 TGGCGCTGCTCCAGATGTCCTCAAGAGCGCTGAACAGACCCAGACCCAGCCG 195
QY 64 AlalleLeuAlaThrSerValAlaIleTyrArgLeu---LeuGlnGluLysGlyTyrGln 82
DB 844 TTGCTGCTGCTTAATTAATAAATAATAATAGATGTTAAGTTTAACATCAATCAAA 897
```

196 GCATCTCTTACGGTTTCGATCCGCGCTCTGGCGCGCTCTGGCTGGCGGAGGGGGTGGCGCGC 355

QY 83 ProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSerGlyAla 102

Db 256 CGCGCGCTTCGTCCCGGGGCACAGCCTGGCGGAATATTCCGCGCTGGTTCGGCGCGGAAGC 315

QY 103 LeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluGluAla 122

Db 316 CTGCGCTTCGCGATGCGGTCAAGCTGTGTGAGCGGTAGGGCCAACTGATGTCAGCAGGGCG 375

QY 123 AlaProAlaAspSerGlyLysMetValAlaValLeuLeuAsnThrProValGluValIleGlu 142

Db 376 GTTCGGCGGGCAGCGCGCATGCCCGATCTCTGGCCTGGGAAGCGCGATGTATTG 435

QY 143 GluAlaCysGlnLysAlaSerGlnLeuGlyValValThrProAlaAsnTyrAsnThrPro 162

Db 436 GCGGCTGTGCGCAGCGCGCCAGGGCGAGGTGTCTAGCGCGGTCAACTTCAACGGCGCG 495

QY 163 AlaGlnIleValIleAlaGlyGluValValAlaValAlaAspArgAlaValGluLeuGln 182

Db 496 GGGCAGGTAGTGTGCGCGTCCCGCGCTGCGTTGAGCGTGCATCGAGGCATCGAAG 555

QY 183 GluAlaGlyAlaLysArgLeuLeuProLeuLysValSerGlyProPheHisThrAlaLeu 202

Db 556 GCACGGCGGCCAACGGCGGTGCGGTTCAGTCAGCGTCCGTCGCATTCGGAATG 615

QY 203 LeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPhe 222

Db 616 ATGCGTCCGCGCCGAGAGTCGCCCGCTCGGTGCGTGGAAAGCTGCAGTCGAGCGCGCG 675

QY 223 ThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAsp---IleAla 241

Db 676 AAGATTTCGCTGTCAGAACGTCAGCGCGCGCGTCCGCGCTGATCTCGATACGCTGCGC 735

QY 242 GlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMet 261

Db 736 CGCGACCTCTCGCACAGCTGTACAGCCCGGTTCGCTGGTGGAGAGCATCCAGCTGCTG 795

QY 262 GlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGlyPhe 281

Db 796 GCGGAAGAAGGGCGTCACGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 855

QY 282 ValLysLys 284

Db 856 AACAGGCGC 864

RESULT 12

US-08-781-986A-25

Sequence 25, Application US/08781986A

Publication No. US2003005436A1

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5255

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/781,986A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

QY 238 GluAspIleAlaGlnLeuThrArgGlnValIysGluProValArgPheTyrGluSer 257
Db 4100 GAAGTAATTAATCAATGTCAGCAATTAATATACAGTACAACTCA 4159
QY 258 IleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysVal 277
Db 4160 ACAGATGGCTAATAGACCAAGGTGTCATCATTTATTGNAATGTCCTGGAAGATT 4219
QY 278 LeuSerGlyPheValIysGlnThrAlaHisLeuAlaHisValGlu 295
Db 4220 TTATCTGGCTTAATATAAAAAAATAATAGATGTTAAGTTAATCAATCA 4273

RESULT 13
US-09-070-927A-458
; Sequence 458, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Steven Barash
; Patrick J. Dillon
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 382
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 458:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6021 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 458:
US-09-070-927A-458

Alignment Scores:
Pred. No.: 9.59e-41 Length: 6021
Score: 421.00 Matches: 84
Percent Similarity: 70.86% Conservative: 23
Best Local Similarity: 55.63% Mismatches: 44
Query Match: 27.73% Indels: 0
DB: 10 Gaps: 0

US-09-308-397-2 (1-306) x US-09-070-927A-458 (1-6021)

QY 154 ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleAlaGlyGluValAla 173

Db 173 GTGGGTCACAGCAAAATTATWATWCMCMCAMCAAAATCGTGAATGGTGGAGGTTCTGCT 232
QY 174 ValAspArgAlaValGluLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLys 193
Db 233 GTTGATCAAGCGATCACACTTCTCAAGAGCGTGGTGTGAAGCGAATGATTCGTTAAAT 292
QY 194 ValSerGlyProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThr 213
Db 293 GTGATGCGCCCTTCCATACGCGCTGTACAACAGCATCAAAAAAATGGCTCAGGAT 352
QY 214 LeuAlaGlnValSerPheSerAspPheThrCysProLeuValGlyAsnThrGluAlaAla 233
Db 353 TTAGCAAAATGCACTTTCACACGATGCAAAATCTCTGTCATTAGTAATACGACTCCCGAA 412
QY 234 ValMetGlnLysGluAspIleAlaGlnLeuLeuThrArgGlnValLysGluProValArg 253
Db 413 ATTATGCCCCAAGAGGCAATTCAGCGTTATTGAAAAAGCAAGTCATGCTCGCGTACGT 472
QY 254 PheTyrGluSerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGly 273
Db 473 TTGAGACAGTATCGAACGATGAGGCTATGACGTAGGACGATGATGAACTGATGAAATG 532
QY 274 ProGlyLysValLeuSerGlyPheValIysLysIleAspGlnThrAlaHisLeuAlaHis 293
Db 533 CCAGGGAACCAATTAATCTGTTTGTAAAAAATTGACAAAAACAATGAAATGCACCGT 592
QY 294 ValGluAspGlnAlaSerLeuValAlaLeuLeu 304
Db 593 GTGGAAGATGTTGCAACATTAACAGAAACGTTA 625

RESULT 14
US-09-815-242-7193
; Sequence 7193, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7193
; LENGTH: 930
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(930)
US-09-815-242-7193

Alignment Scores: 2,18e-41 Length: 930
 Pred. No.: 416.50 Matches: 106
 Score: 416.50
 Percent Similarity: 51.80% Conservative: 52
 Best Local Similarity: 34.75% Mismatches: 138
 Query Match: 27.44% Indels: 9
 DB: 10 Gaps: 5

US-09-308-397-2 (1-306) x US-09-815-242-7193 (1-930)

QY 5 AlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyrAsp 24
 Db 10 GCGCTATTATTTCCAGGCAAGGCTCGCAATGTAGGAATGGGAAATCATTTCTATGAG 69
 QY 25 GlnTyrProIleValIysGlnThrIleAspArgAlaSerGlnValLeuGlyTyrAspLeu 44
 Db 70 AGCCACACTCTAGCTAAAGAAATGTTTGAAGGGCTTTAAGCCACTTAAGTGGATAG 129
 QY 45 ArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyrThrGlnProAla 64
 Db 130 AAAAAAACGCTTTTGAAGAAATGAGCTTTTAAAAAGAGAGCGCTTACACCCAGCGCTGCC 189
 QY 65 IleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLys-----GlyTyr 81
 Db 190 ATTATTAGTGAGCTATATCGCTTACCNAATTGCTCAACAGCAAGTAATAGGGGGTTA 249
 QY 82 GlnProAspMetValAlaGlyLeuSerLeuGlyGlnTyrSerAlaLeuValAlaSerGly 101
 Db 250 AAACCGGTTTTCCTTAGCGCATTCGCTCGCGAAGTGAGCGCGTCTCTTTGAGTGGG 309
 QY 102 AlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluGlu 121
 Db 310 GCGTTAGATTGTAAGAAAGCCCTTAAACTCAGCCACCAAGAGGCAAAATGATCAAGAA 369
 QY 122 AlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIle 141
 Db 370 GCGTCGCGAATAAAGACGCTTCCATGATGTCGTTTGGCGTTCTTGAAGAAGCGCTT 429
 QY 142 GluGluAlaCysGlnLysAlaSerGlnLeuGlyValValThrProAlaAsnTyrAsnThr 161
 Db 430 TTGAGTTGTGTCAAGAACACCAAAAT-----GTGTGTGCGCGAATTTCAATGGC 480
 QY 162 ProAlaGlnIleValIleAlaGlyGluValValAlaValAlaValArgAlaValGluLeu 181
 Db 481 GSCATGCAAGTGGTTTATAGCGGGATTAAAGACGATTGAAAGCCCTAGACCCGACTTA 540
 QY 182 GlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAla 201
 Db 541 AAGGAAATGGGGCTAAAAAGAGTGTGTTTTTATAGAAATGAGCGTGGCGGAGCCATG 600
 QY 202 LeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAsp 221
 Db 601 TTTTGTAGAGCCTATGACTTTTAAATTCAGGAATTTGCTAGAAATAAGCCCTGAAAGATAA 660
 QY 222 PheThrCysProLeuValGlyAsn---ThrGluAlaAlaValMetGlnLysGluAspIle 240
 Db 661 TTCCATTTTGAATCATCTCCATCGACTAATGAAGCGTATCAACAAAGCAAAAGGCC 720
 QY 241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
 Db 721 GTTGAACATTGAGGTTCAACTACTCAGCGGGTGGCTTATCAAGACTGCGTGAATGCC 780
 QY 261 MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280
 Db 781 AACATGACCGA---GTGGATGCTCTTTTGTGATTTAGCTGTGGAGTGTGTTAAAGGGG 837
 QY 281 PheValLysIleAspGlnThrAlaHisIleAlaHisValGluAspGlnAlaSerLeu 300
 Db 838 CTTAAACAGCGATTAAAGCAACAAACCCACCATTAAGC---GTGGGGGATAATAAGGGGCTT 894
 QY 301 ValAlaLeuLeuGlu 305
 Db 895 GATGAGCGCATTTGAG 909

RESULT 15
 US-09-974-300-5552
 ; Sequence 5552, Application US/09974300
 ; Patent No. US20020146721A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Berka, Randy M.
 ; APPLICANT: Clausen, Ib Groth
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
 ; TITLE OF INVENTION: Expression
 ; FILE REFERENCE: 10085.500-US
 ; CURRENT APPLICATION NUMBER: US/09/974,300
 ; PRIOR FILING DATE: 2001-10-05
 ; PRIOR APPLICATION NUMBER: 09/680,598
 ; PRIOR FILING DATE: 2000-10-06
 ; PRIOR APPLICATION NUMBER: 60/279,526
 ; PRIOR FILING DATE: 2001-03-27
 ; NUMBER OF SEQ ID NOS: 8481
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5552
 ; LENGTH: 592
 ; TYPE: DNA
 ; ORGANISM: Bacillus clausii
 ; US-09-974-300-5552

Alignment Scores: 1.29e-41 Length: 592
 Pred. No.: 416.00 Matches: 93
 Score: 416.00
 Percent Similarity: 63.59% Conservative: 31
 Best Local Similarity: 47.69% Mismatches: 69
 Query Match: 27.40% Indels: 2
 DB: 10 Gaps: 2

US-09-308-397-2 (1-306) x US-09-974-300-5552 (1-592)

QY 3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22
 Db 7 AAGTTGCGCTTTTATTTCTGACAGAGGCTCACAAAAAGTTGCGATGGGCAACGCTT 66
 QY 23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
 Db 67 TTGCTGAAGATTGTATGTCACAAAGGAAGTCTTTGAGCGGCTGACCAAGCGCTTGGTTTT 126
 QY 43 AspLeuArgTyrLeuIleAsp---ThrGluGluAspLysLeuAsnGlnThrArgTyrThr 61
 Db 127 CCTTTTCAGAGATCATCGAACCATGACAGCAAGCAAGCTGAACAAACGGCTTTATGCC 186
 QY 62 GlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGlyTyr 81
 Db 187 CAACCTGCTTTAGTGACGATGAGCACTGCAGTCTCTCAATTGTTCCGCGACGCTGACATT 246
 QY 82 GlnProAspMetValAlaGlyLeuSerLeuGlyGlnTyrSerAlaLeuValAlaSerGly 101
 Db 247 CAAGCAGATTTTTCGAGGCCATAGCTAGGGGAGTATTCGCTTTTATGGCATGTAAG 306
 QY 102 AlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluGlu 121
 Db 307 TCCTTACATTGAAGATGGGTTCACATTTGTCATCAACGGGGGACACTCATGGAGGAA 366
 QY 122 AlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIle 141
 Db 367 GCTGTGCCCAAGCAAGGCGGATGGCAGCGGTACTCGGCTTAATAAAGAGGAGCTT 426
 QY 142 GluGlu---AlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
 Db 427 GAGGAATGGCTTCCGAATTTGAGCGGAGCGGGAAGTTGCTGAACCTTCCCAATTTGANT 486
 QY 161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAlaAspArgAlaValGluLeu 180
 Db 487 TGTCCAGGGCAAAATTCCTGTCTGAGGAGTGGGAAAGGAATGAACACAGCAGCAGTATTA 546
 QY 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSer 195

Db 547 GCGAAACAAAAGGGGCCAAGCGTGTCTTTGCCACTTGGCGTCAGC 591

Search completed: June 11, 2003, 22:27:08
Job time : 774 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 11, 2003, 20:10:52 ; Search time 1430 Seconds
(without alignments)
3465.608 Million cell updates/sec

Title: US-09-308-397-2

Perfect score: 1518

Sequence: 1 MTKTAFAGGQAYLGMGR.....QTAHLAHVEDQASLVALLEK 306

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09308397/runat_06062003_112349_28885/app_query.fasta_1.455
-DB=EST -QMT=fastap -SUFFIX=rat -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NCR=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09308397@cgn_1_1525 @runat_06062003_112349_28885 -NCPU=6 -ICPU=3
-NO_MMAB -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_estc.*
9: gb_esti.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_man.*
24: em_gss_mus.*
25: em_gss_oth.*
26: em_gss_pro.*
27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	327	21.5	787	12	BG584476 EST486236
2	324.5	21.4	511	14	BQ818656 1030072D0
3	314.5	20.7	544	9	AU069821 AU069821
4	296.5	19.5	647	9	AJ500905 AJ500905
5	293.5	19.3	690	9	AJ498681 AJ498681
6	291	19.2	647	12	BG851744 1024032B1
7	290	19.1	585	10	BE404963 WHE1207.B
8	285.5	19.1	491	12	BF512296 UI-H-BW1-
9	278.5	18.3	648	10	BE440632 sp50b03.Y
10	278	18.3	539	13	BJ462718 BJ462718
11	274.5	18.1	850	12	BG444706 GA_Ea002
12	274	18.1	513	13	BJ462722 BJ462722
13	270	17.8	892	11	AV108033 Zea mays
14	270	17.8	1237	17	AF029428 AF029428
15	269	17.7	599	10	BE237947 BE237947
16	269	17.7	708	12	BG441502 GA_Ea001
17	267	17.6	659	12	BE725078 894080H05
18	266	17.5	684	12	BG441246 GA_Ea001
19	265.5	17.5	809	12	BG418889 HVSMEK002
20	264	17.4	658	12	BG591410 EST499252
21	264	17.4	744	12	BG597049 EST495727
22	263.5	17.4	827	13	BM359182 GA_Ea001
23	263	17.3	648	9	AJ500837 AJ500837
24	260.5	17.2	602	13	BM525587 sal29g11.
25	258	17.0	587	13	BM524467 sal16a10.
26	257	16.9	742	13	BI934864 EST554753
27	256	16.9	582	10	BE43678 WHE1121.E
28	253	16.7	816	12	BF630807 HVSMEK001
29	252.5	16.6	585	13	BJ344706 BJ344706
30	252	16.6	1104	12	BE795116 601592648
31	252	16.6	1106	12	BF789052 602104910
32	250.5	16.5	613	10	AV926272 AV926272
33	249	16.4	582	10	AM617993 EST314067
34	248	16.3	668	12	BG859575 1024064B1
35	247.5	16.3	604	10	AV974832 AV974832
36	246	16.2	755	13	BI919675 BI919675
37	246	16.2	788	13	BI436174 EST538935
38	244	16.1	628	12	BG846299 1024013B0
39	243	16.0	577	10	AV924867 AV924867
40	243	16.0	601	13	BI064433 Pgfln.pk0
41	242	15.9	642	14	BQ462913 H102G11r
42	241	15.9	796	12	BG699020 602678651
43	239.5	15.8	534	9	AJ498573 AJ498573
44	239	15.7	575	13	BI500434 rs72612.Y
45	239	15.7	699	9	AL508847 AL508847

ALIGNMENTS

RESULT 1
BG584476

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

BG584476 787 bp mRNA linear EST 11-APR-2001
EST486236 MHAM Medicago truncatula/Gloms versiforme mixed EST
library CDNA clone pMHAM-15P1 5' end, mRNA sequence.

BG584476
BG584476.1 GI:13599540

EST
Medicago truncatula/Gloms versiforme mixed EST library.
Medicago truncatula/Gloms versiforme mixed EST library
Eukaryota; mixed EST libraries.

1 (bases 1 to 787)
Harrison M.J., Liu J., Town C.D., Van Aken S., Utterback T., Cho J.
and Fraser C.M.

ESTs from roots of Medicago truncatula after colonization with

Glomus versiforme, 2001
 Unpublished (2001)
 Contact: Harrison M.J.
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73401
 Tel: 580-223-5810
 Fax: 580-221-7380
 Email: mjharrison@noble.org
 Noble EST name: N379946e TIGR sequence name: MTDB085TK More
 information is available at: <http://www.medicago.org>
 Seq primer: SKmod (CTA gaa gta gat cc).
 Location/Qualifiers
 1..787
 /organism="Medicago truncatula/Glomus versiforme mixed EST
 library"
 /cultivar="Medicago truncatula genotype Al7"
 /db_xref="taxon:119092"
 /clone="PMHAM-15P1"
 /clone_lib="PMHAM"
 /tissue_type="roots colonized with Glomus versiforme"
 /dev_stage="roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Glomus versiforme. The library was
 made from a mixture of RNA from each of these stages."
 /lab_host="E. coli strain XL0LR"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Glomus versiforme. The cDNA was
 directionally ligated into the Unizap XR vector from
 Stragene and packaged using gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assist
 helper phage and propagated in XL0LR cells."
 BASE COUNT 204 a 154 c 206 g 223 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,93e-31 Length: 787
 Score: 327.00 Matches: 80
 Percent Similarity: 55.30% Conservative: 40
 Best Local Similarity: 36.87% Mismatches: 85
 Query Match: 21.54% Indels: 12
 DB: 12 Gaps: 5
 US-09-308-397-2 (1-306) x BG584476 (1-787)
 QY 5 AlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyrAsp 24
 |||||
 DB 97 GCATTCTCTCCCTGGCAGGTGCACAAAGCTGTGGAATGGGAAAGAA---GCTCAA 153
 |||||
 QY 25 GlnTyrProIleVallysgluThrIleAspArgAlaSerGlnValLeuGlyTyrAsp--- 43
 |||||
 DB 154 AGTGTCTCTGCTGCTGCATCTTTGTACAAAGAACCAATGAGATTTCGGGGTTGATCTT 213
 |||||
 QY 44 LeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyrThrGlnPro 63
 |||||
 DB 214 CTTGATGATGATTATGAGCCCAAGGAAAGAACTGAATTCACAGTTATTAGTCAGCT 273
 |||||
 QY 64 AlaIleLeuAlaThrSerValIleIleTyrArgLeuLeu-----GlnGluLysGlyTyr 81
 |||||
 DB 274 GCCTGTATGTCACAAAGTCTTGCTGTGTGTAGCTACTTCGGGCACGTGAGGAGGTGAG 333
 |||||
 QY 82 Gln-----ProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeu 97
 |||||
 DB 334 CAGATTATTGATCTCTTGATGTTACATCGCGTTTGAGCTGGAGATATACTCTGCTG 393
 |||||
 QY 98 ValAlaSerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAla 117
 |||||
 DB 394 GCATTTCGCTGGGCTTCACGCTTTGAAGATGACATGAAATTCGTGAAACTGAGGGGTGAA 453
 |||||
 QY 116 TyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrPro 137
 |||||

Db 454 GCCATGCAGATGCTTCTGATGCTGTATAAAGTGCATGTTAGTGTGGTAGATTGGAC 513
 QY 138 ValGluValIleGluGluAlaCys-----GlnLysAlaSerGlnLeuGlyVal 153
 |||||
 Db 514 TCAGAAAAAGTCCAGCAGTTGTGTGATGCGAGCAATCAGGAAGTGCCTGAAGCTGAGAAG 573
 |||||
 QY 154 ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleAlaGlyGluValValAla 173
 |||||
 Db 574 GTTCAGATTGCCAATTACCTATGTCGAGGAACACTATGCTCTCTGAGGCATAAAGGA 633
 |||||
 QY 174 ValAspArgAlaValGluLeuLeuGlnGluAlaGlyAlaLysArgLeuLeuProLeuLys 193
 |||||
 Db 634 ATAGAAGCGGTGGAAGCCAAAGTCTTTCAAGGCTCGAATGACTGTGCGCTTAGCT 693
 |||||
 QY 194 ValSerGlyProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeu 210
 |||||
 Db 694 GTTGGCGGAGCTTTCATCTACTAGTTTATGGAACCTGCGGTGTCAGAGATTG 744
 |||||
 RESULT 2
 BQ818656/c 511 bp mRNA linear EST 01-AUG-2002
 LOCUS
 DEFINITION 1030072D02.y1 C. reinhardtii CC-1690, Deflagellation (normalized),
 Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
 ACCESSION BQ818656 GI:22068897
 VERSION BQ818656.1
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadales; Chlamydomonas.
 REFERENCE 1 (bases 1 to 511)
 AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre
 P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.
 TITLE Analysis of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants. Project: 1030
 JOURNAL Unpublished (2002)
 COMMENT Contact: Charles Hauser
 DCMB Box 91000
 Duke University
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu.
 FEATURES
 Location/Qualifiers
 1..511
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, Deflagellation
 (normalized), Lambda Zap II"
 /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
 XhoI; Deflagellation library, constructed by John Davies
 and Jeffrey McDermott, combines cDNAs from CC-1690 cells
 which had been re-synthesizing flagella for 15, 30 and 60
 min after being deflagellated by pH shock. PolyA mRNA was
 purified from each sample, pooled and cDNA synthesized.
 The cDNA was directionally cloned into lambda Zap II
 (Stragene) in the EcoRI (5') and XhoI (3') sites.
 pBluescript II SK- plasmids were excised from the lambda
 Zap clones by superinfection with EXassist (Stragene)
 phage. The library was normalized using method 4 described
 in Bonaldo et al., (1996) Genome Research 6: 791-806."
 BASE COUNT 107 a 148 c 126 g 130 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,93e-31 Length: 511
 Score: 324.50 Matches: 69
 Percent Similarity: 63.25% Conservative: 36
 Best Local Similarity: 41.57% Mismatches: 60
 Query Match: 21.38% Indels: 1
 DB: 14 Gaps: 1

```

US-09-308-397-2 (1-306) x B0818656 (1-511)
QY 141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProLalaSerThrAsn 160
Db 511 ATTGCGAAGCGGTGAAGAAGCTGCAGAGGTCAGTCTCTCCGCTAAACTTTAAC 452
QY 161 ThrProLalaGlnLeuValIleAlaGlyGluValValAlaValAlaValAlaValGluLeu 180
Db 451 TCTCCGGGACAGGTGGTTATTGCGGTATTAAGAGACGGTTCAGCGTCTGCGCGTCC 392
QY 181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr 200
Db 391 TGTAAGCGCGGGCGGCAAAACCGCGCTGCCCTTACCAGTGAGCGTACCTCTACTCT 332
QY 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
Db 331 GCGCTGATGAACACGACGCGCAAACTGGCAGTAGAATTAGCGAAATACCTTTAAC 272
QY 221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAsp--- 239
Db 271 GCACCAACAGTCTCTGTTGTGAATAGTTGTAATGCAATGCCAAACCAATGGTGATGCC 212
QY 240 IleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheThrGluSerIleGly 259
Db 211 ATCCGTGACCACTGGTACGTACGTCTGTTGATAAACCCTGTCAGTGACGAAAGTCTGTGAG 152
QY 260 ValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSer 279
Db 151 TACATGCGACGCGCAAGCGTAGACATCTCTATGAAGTCGCGCGCGGCAAGTCTACT 92
QY 280 GlyPheValLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSer 299
Db 91 GGCCTGACGAAACGCAATGTGACACACCTGACCGCTCGCGCTGAAACGAACTTCAGCG 32
QY 300 LeuValAlaLeuLeuGlu 305
Db 31 ATGCGACGCGCGCTCGAG 14

RESULT 3
AUG069821/c
LOCUS
DEFINITION AUG069821 Rice panicle at flowering stage Oryza sativa (japonica cultivar-group) cDNA clone E3726_7A, mRNA sequence.
ACCESSION AUG069821
VERSION AUG069821.1 GI:5004646
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group).
ORGANISM Oryza sativa (japonica cultivar-group).
REFERENCE
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from panicle at flowering stage
JOURNAL Unpublished (1996)
COMMENT National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: teasaki@abrr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
PROJECT = 'RGP',
Location/Qualifiers
source 1..544
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db xref="taxon:39947"
/clone="E3726_7A"
/clone.lib="Rice panicle at flowering stage"
/dev stage="flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 118 a 161 c 126 g 133 t 6 others
ORIGIN
Alignment Scores:
Pred. No.: 4,26e-30 Length: 544
Score: 314.50 Matches: 79
Percent Similarity: 61.54% Conservative: 33
Best Local Similarity: 43.41% Mismatches: 67
Query Match: 20.72% Indels: 4
DB: 9 Gaps: 1
US-09-308-397-2 (1-306) x AUG069821 (1-544)
QY 107 AspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluGluAlaAlaProAlaAsp 126
Db 541 GATCGGTGCTCTGGTTGAGATCGCGGAGTTCATGANGAAGCGCTACCGGAGGC 482
QY 127 SerGlyLysMetValAlaValLeuAsnThrProValGluVal-IleGluGluAlaCysGI 146
Db 481 ACGGCGCTATGCGGCAATCATCGTCTGAATGATCGTCTATTGCGAAGCGTGTNA 422
QY 146 nLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThrProAlaGlnIleVa 166
Db 421 AGAAGCTCAGAGAGGTACGGTCTCTCCGCTAACTTTAACTCTCCGGA-CAGGTGGT 363
QY 166 IleAlaGlyGluValValAlaValAlaValAspArgAlaValGluLeuGlnGluAlaGlyAl 186
Db 362 TATTGCGGTCTATAAAGAGCGGTGAGCGTCTGCGCTGTAAGCGCGGCGGC 303
QY 186 a-LysArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeuLeuGluProA 206
Db 302 NAAACGCGCGCTCCCTTACAGTGACGTACCGTCTCTACTGTGCGCTGATGAACACG 243
QY 206 LaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPheThrCysProL 226
Db 242 CAGCGCAAACTGGCAGTAGAATTAGCGAAATCACCTTTAACGCCAACACAGTCTCTG 183
QY 226 euValGlyAsnThrGluAlaAlaValMetGlnLysGluAsp---IleAlaGlnLeuLeuT 245
Db 182 TTGTGAATAACGTTGATGTGAATGCGAAACCAATGGTGATGCCATCCGTACCACTGG 123
QY 245 hrArgGlnValLysGluProValArgPheThrGluSerIleGlyValMetGlnGluAlaG 265
Db 122 TACGTCACTGTATACCCGGTTTCAGTGAGCAAGTCTGTGTAGTACATGACGCGCAAG 63
QY 265 lylleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGlyPheValLysI 285
Db 62 GCGTAGAACATCTCTATGAAGTCGCGCGGCAAAAGTCTTACTGGCTGACGAAACGCA 3
QY 285 le 285
Db 2 TT 1

RESULT 4
AUG500905
LOCUS
DEFINITION AUG500905 MTGIM Medicago truncatula cDNA clone mtgmaccl2002le04, mRNA sequence.
ACCESSION AUG500905
VERSION AUG500905.1 GI:22081838
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
REFERENCE
AUTHORS Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
1 (bases 1 to 647)
Krajinski,F., Manthey,K., Bartelsmeier,V., Meyer,F., Bartels,D., Bekel,T., Linke,B., Franken,P., Kuester,H., Perlick,A.M. and Puhler,A.
Detection of transcript sequences from mycorrhizal roots of the

```



```

Db      131 ACATGCGGTTGACCTGGGAGATATACGCTCTGCGCATTTGCTGGGGCTTTTCAGCTTT 190
Qy      106 GluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluGluAlaAlaProAla 125
Db      191 GAAGATGGACTGAATTTGGTGAACACTGAGGGGTGAAGCCATGAGGATGCTTCTGATGCT 250
Qy      126 AspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGluGluAlaCys 145
Db      251 GTTAAAGTGCATGGTTAGTGTGGTAGGATTGGACTCAGAAAAGTCCAGCAGTTGTGT 310
Qy      146 -----GlnLysAlaSerGluLeuGlyValValThrProAlaAsnThr 161
Db      311 GATCAGCGAATCAGGAAGTGCCTGAAGCTGAGAAGTTCCAGATTGCCAATTACCTATGT 370
Qy      162 ProAlaGlnIleValIleAlaGlyGluValValAlaValAlaAspArgAlaValGluLeuLeu 181
Db      371 CCAGAAACTATGCTGCTCTGGAGGCATAAAGGAATAGAAGCGGTGGNAGCAAGCA 430
Qy      182 GlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAla 201
Db      431 AAGTCTTTCAAGGCTCAATGACTGTGCGCTTAGCTGTTCGGGAGCTTTCCATACTAGT 490
Qy      202 LeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAsp 221
Db      491 TTAAGAACTGCGGTGTCAGATTGGAAGCTGCATTTGCAATGCGACAGACATAGAACC 550
Qy      222 PheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAsp---Ile 240
Db      551 CCAAGATACACAGTCATTTCCATGTGAATTCCTGCGGCACACAGATCCTGACACAATA 610
Qy      241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
Db      611 AAGAGGATATCGGCACACAGGTTACTTCCCTGTTCAATGGGAAACAACCTGTGAAGACT 670
Qy      261 MetGlnGluAlaGlyIle 266
Db      671 CTTCTATCCAAAGGGGCTG 688

RESULT 6
BG851744
LOCUS      1024032B11.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION  BG851744
VERSION     BG851744.1 GI:14232928
KEYWORDS   EST.
SOURCE      Chlamydomonas reinhardtii.
ORGANISM    Chlamydomonas reinhardtii.
REFERENCE   1 (bases 1 to 647)
AUTHORS     Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
Mcdermott,J.P., Sliflow,C., Stern,D. and Surzycki,R.
TITLE       Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
JOURNAL     Unpublished (2000)
COMMENT     Contact: Charles Hauser
            DCM5 Box 91000
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
            Location/Qualifiers
            1. .647
            /organism="Chlamydomonas reinhardtii"
            /strain="CC-1690 wild type mt+ 21gr"
            /db_xref="taxon:3055"
            /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
            II"
            /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:

```

XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO₂ and HS medium bubbled with 5% CO₂. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 128 a 177 c 214 g 128 t
ORIGIN

Alignment Scores:
Pred. No.: 5,42e-27 Length: 647
Score: 291.00 Matches: 69
Percent Similarity: 58.72% Conservative: 32
Best Local Similarity: 40.12% Mismatches: 61
Query Match: 19.17% Indels: 10
DB: 12 Gaps: 5

US-09-308-397-2 (1-306) x BG851744 (1-647)

```

Qy      7 LeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyrAspGlnTyr 26
Db      11 TTGTTCCCGCCAGGTCGCGACAGCGTCGCGATGGCCAAAGGACCTTGTGGCCACCGTG 70
Qy      27 ProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyrAsp---LeuArg 45
Db      71 CCGAAGGCCAAGGAGATGTTGCAAGCGCTCTGAGATCCTGGGCTACGACCTGCTGAAG 130
Qy      46 TyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyrThrGlnProAlaIle 65
Db      131 GTGTCGTGGAGGGCCCAAGGAGAAGCTGCACAGCACCGCGGTGAGTACGCCCGCTATC 190
Qy      66 ---LeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGlyLysGlyTyrGlnPro--- 83
Db      191 TACGTGGCTCGCTGGCGCGCTGGAGAAGCTGCGCGCGAGGCGGCGCCGCCCAT 250
Qy      84 -----AspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
Db      251 GACGCCATTGACGTGGCGTGGCGCTGTCTGCTGGCGAGTACACGGCCCTGGCTCGCT 310
Qy      101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
Db      311 GGTGCATGAGCTTCGAGGACGGCTGCGGTGTGAAGCTGCGCGCGAGTCCATGCAG 370
Qy      121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
Db      371 GCGGCTGCTGACGCGCAGCCAGCAGCATGTGTCTTCCTCATTTGGGTGGAGCGGCCAAG 430
Qy      141 IleGluGluAlaCysGlnLysAlaSerGlu---LeuGlyValValThrPro----- 156
Db      431 GTGGCGAGCTGTGCAAGTCCGCCAGCAGCAGTGGTGTGAGGAAAGCGCCGCTCAGATC 490
Qy      157 AlaAsnTyrAsnThrProAlaGlnIleValIleAla 168
Db      491 GCCAACTACCTGTGCCCTGTCTACCTACGCGCGTGTGC 526

```

RESULT 7

```

BG851744
LOCUS      BE404963
DEFINITION WHEI1207_B01_D012S Wheat etiolated seedling root cDNA library
            Triticum aestivum cDNA clone WHEI1207_B01_D01, mRNA sequence.
ACCESSION  BE404963
VERSION     BE404963.1 GI:9364431
KEYWORDS   EST.
SOURCE      bread wheat.
ORGANISM    Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

```

REFERENCE
 AUTHORS Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L., and Tong, J.C.
 TITLE The structure and function of the expressed portion of the wheat genomes
 JOURNAL Unpublished (2000)
 COMMENT Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: candersn@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
 Seq primer: Strategene SK primer.

FEATURES

source
 1..585
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE1207_B01.D01"
 /clone_lib="Wheat etiolated seedling root cDNA library"
 /tissue_type="Root"
 /dev_stages="Five day old etiolated seedling"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Seeds were surface-sterilized, germinated and grown aseptically in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Roots were harvested. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give plasmids in the 10 Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
 BASE COUNT 154 a 121 c 157 g 153 t
 ORIGIN

Alignment Scores:

Pred. No.: 7,27e-27 Length: 585
 Score: 290.00 Matches: 72
 Percent Similarity: 58.19% Conservative: 31
 Best Local Similarity: 40.68% Mismatches: 62
 Query Match: 19.10% Indels: 12
 DB: 10 Gaps: 5

US-09-308-397-2 (1-306) x BE404963 (1-585)

QY 4 ThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyr 23
 |||||
 DB 58 ACCGATTCCTTTCCCGCCAGCGTCTCAGGCTGTGGAATGGGTGAAGAGGCTCTT 117
 |||||
 QY 24 AspGlnTyrProIleValIleGluThrIleAspArgAlaSerGlnValLeuGlyTyrAsp 43
 |||||
 DB 118 AAT---GTTTCAGCAGCTGCAGAACTATTTCATAAGGCAATGATATCTGGTATGAC 174
 |||||
 QY 44 LeuArgTyrLeu---IleAspThrCluGluAspIleValLeuGlnThrArgTyrThrGln 62
 |||||
 DB 175 TTGTTGAATCTTTGATCGATGGACCAAGAAAGCTGAATTCACAGTGATCAGTCAG 234
 |||||
 QY 63 ProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeu-----GlnGlyLeuGly 80
 |||||
 DB 235 CCAGCTATATATGTTTACAGCTTCAGCTAGAGGTGTCGGCGCAGCGGAAGAGGC 294
 |||||
 QY 81 Tyr-----GlnProAspMetValIleGlyLeuSerLeuGlyGluTyrSerAla 96
 |||||
 DB 295 CAATCTGTAATTAATCCGTAGATGTACATGTGCTCTCAGCTGGGAGATATACCGCG 354
 |||||

QY 97 LeuValAlaSerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGly 116
 |||||
 DB 355 CTTGCAATTTGGTGGTGCCTTTAGCTTTGAGGATGCTGCAAGCTTTGTCAAGCTTAGAGGA 414
 |||||
 QY 117 AlaTyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThr 136
 |||||
 DB 415 GAAGCATGACAGATGCCCTCAGACGCTGCCAATATGCGATGGTGTAGTGTATTTGGTCTA 474
 |||||
 QY 137 ProValGluValIleGluGluAlaCysGlnLysAlaSer-----GluLeuGly 152
 |||||
 DB 475 GATTTCAGAAAGGTGCACCAATATTCGATGTCGAAATGAGGATGTCGATGAAAGGAA 534
 |||||
 QY 153 ValValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleAlaGly 169
 |||||
 DB 535 AGAGTTCAATAGCAAAATTTCTCTGCTCTGTAACCTATGCAAGTTTCGTT 585
 |||||
 BF512296 491 bp mRNA linear EST 07-DEC-2000
 UI-H-BM1-amb-e-04-0-UI.sl.NCI.CGAP_Sub7 Homo sapiens cDNA clone
 IMAGE:3069391 3', mRNA sequence.
 BF512296
 BF512296.1 GI:11597475
 EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 491)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaepsr@mail.nih.gov
 Oligo-dT track not found, Not 1 site shown in beginning of sequence
 is likely internal to the message. cDNA Library Preparation: M.B.
 Soares Lab Clone distribution: NCI-CGAP clone distribution
 information can be found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: M13 Forward
 POLYA=No. Location/Qualifiers
 1..491
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3069391"
 /clone_lib="NCI CGAP Sub7"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pTT3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not 1; Site 2: Eco RI; NCI-CGAP_Sub7
 is a subtracted library derived from NCI-CGAP_Sub6. The
 NCI-CGAP_Sub7 library had 12 million recombinants. A
 single-stranded DNA preparation of NCI-CGAP_Sub6 was used
 as a tracer in a subtractive hybridization with a driver
 comprising the IMAGE pool (NCI-CGAP_Kid3 pool 1 LLAM
 3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)
 1322376-1323911, 1456008-1456775, 1500552-1502855);
 (IMAGE Clones 1323912-1325831, 1471368-1472903,
 1492104-1493255); NCI CGAP_Lu5 pool 1 LLAM 3575-3582,
 3851-3854 (IMAGE Clones 1414920-1417991, 1520904-1522439
); NCI CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720,
 3733-3735 (IMAGE Clones 1257096-1258631, 1469064-1470983
 , 1475592-1476743); NCI CGAP_Pr22 pool 1 LLAM 2457-2459,
 2758-2759, 3062-3068 (IMAGE Clones 985608-986759,
 1101192-1101959, 1217928-1220615); NCI CGAP_Co10 pool 1
 LLAM 2644-2653, 2871-2872 (IMAGE Clones 1057416-1061255
 , 1144584-1145351). (6% of the driver population), plus a
 pool of 3,840 arrayed clones from NCI-CGAP_Sub1 (IMAGE
 Clones 2708616-2710535) and NCI-CGAP_Sub2 (IMAGE
 Clones 2710536-2712455) (4% of the driver population
), plus a pool of 11,136 clones from NCI-CGAP_Sub3 (IMAGE


```

Db      256 GTCTCGGGGATTA-----AAAGCAGTAGAAGTCTGGAATCCAGCAAG 303
QY      186 -----AlaIyAargLeuIleProLeuIySValSerGlyPropheHisThrAlaLeu 202
Db      304 TCTTTCAAGCTCGAATGACGTGGCGCTAGCTGTGCTGTGCTTTCCACACTAGTTT 363
QY      203 LeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPhe 222
Db      364 ATGAACACAGCTGTGTCAAGGTTGAAGCAGCATTTGGCAACCAACAGAAATCAGAACCCCA 423
QY      223 ThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAsp---IleAla 241
Db      424 AGAATACCATGTCATCTCCATGTGATGCACAGCCACATACAGATCCCTGATCTCAATAAG 483
QY      242 GlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMet 261
Db      484 AAGATATTGGCATGCCANGTTACCTCACCTGTTCAATGGGAAACCAACAGTCAGACTCTT 543
QY      262 GlnGluAlaGlyIleSerAsnThrIleGluIleGlyProGlyLysValLeuSerGlyPhe 281
Db      544 CTAACCAAGGCTCTGAAGAAAGCTATGAATCTGGACCCCGGAAGGTTATTGCTGGAAATC 603

RESULT 10
BU462718      539 bp      mRNA      linear      EST 23-MAY-2002
LOCUS      BU462718 K. Sato unpublished cDNA library, cv. Haruna Nijo
DEFINITION      germination shoots Hordeum vulgare subsp. vulgare cDNA clone
                bag924b06 5', mRNA sequence.
ACCESSION      BU462718
VERSION      BU462718.1 GI:21141225
KEYWORDS
SOURCE      Hordeum vulgare subsp. vulgare.
ORGANISM      Hordeum vulgare subsp. vulgare.
REFERENCE      1. (bases 1 to 539)
AUTHORS      Sato,K., Saisho,D. and Takeda,K.
TITLE      Barley EST sequencing project in NIG and Okayama Univ
JOURNAL      Unpublished (2002)
COMMENT      Contact: Tadasu Shin-i
                Center For Genetic Resource Information
                National Institute of Genetics
                1111 Yata, Mishima, Shizuoka 411-8540, Japan
                Tel: 81-559-81-6856
                Fax: 81-559-81-6855
                Email: tshini@genes.nig.ac.jp.
FEATURES
    source
        1..539
            /organism="Hordeum vulgare subsp. vulgare"
            /cultivar="Haruna Nijo"
            /db_xref="taxon:112509"
            /clone="bag924b06"
            /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
            Nijo germination shoots"
            /tissue_type="shoots"
            /dev_stage="germination"
BASE COUNT      147 a 115 c 139 g 138 t
ORIGIN
Alignment Scores:
    Pred. No.:      2,27e-25      Length:      539
    Score:      278.00      Matches:      66
    Percent Similarity:      60.39%      Conservative:      27
    Best Local Similarity:      42.86%      Mismatches:      53
    Query Match:      18.31%      Indels:      8
    DB:      13      Gaps:      4
US-09-308-397-2 (1-306) x BU462718 (1-539)

```

```

QY      4 ThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyr 23
Db      34 ACCGCGTTCCTTTCCCGCCAGGGTGTCTCAGGCTGTGGATGGGTAAGAAGCTCTT 93
QY      24 AspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyrAsp 43
Db      94 AAT---GTTGAGCAGCTGCAGAACTATTGTAGAGCAATATGATATATTGCTACGAC 150
QY      44 LeuArgTyrLeu---IleAspThrGluGluAspLysLeuAsnGlnThrArgTyrThrGln 62
Db      151 TTCTCAATCTTTGCATCGATCGACCAAAAGAAAGCTCACTCAACAGTCAGTCAG 210
QY      63 ProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeu-----GlnGluLysGly 80
Db      211 CCAGCTATATATGTTACCAGCCTTGCAGCTGTAGAAGTGTCTACGTCACGGGAAGACGGT 270
QY      81 Tyr-----GlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAla 96
Db      271 CAATCTGTAATTAACCTCTGTAGATGTGCATGTGTCTCAGCTTGGGAGATATATACCGCG 330
QY      97 LeuValAlaSerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGly 116
Db      331 CTTCGATTTGCTGTGCTCTTACGCTTTGAGGATGGTCTGAAGCTTGTCAAGCTTAGAGGA 390
QY      117 AlaTyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThr 136
Db      391 GAAGCCATGAGGATGCTCAGATGTGCTCAATAGTCGATGGCTAGTGTGATTGCTCTA 450
QY      137 ProValGluValIleGluGluAlaCysGlnLysAlaSerGlu 150
Db      451 GATTCAAAAAGGTGCAACAATATATCGATGTGCTCAATGAG 492

RESULT 11
BG444706      850 bp      mRNA      linear      EST 15-MAR-2001
LOCUS      BG444706 GA_Ea0025E19f Gossypium arboreum 7-10 dpa fiber library Gossypium
DEFINITION      arboreum cDNA clone GA_Ea0025E19f, mRNA sequence.
ACCESSION      BG444706
VERSION      BG444706.1 GI:13354358
KEYWORDS      EST.
SOURCE      Gossypium arboreum.
ORGANISM      Gossypium arboreum.
REFERENCE      1. (bases 1 to 850)
AUTHORS      Wang,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
                D., Wood,T.C., Leslie,A. and Wilkins,T.A.
TITLE      An integrated analysis of the genetics, development, and evolution
                of the cotton fiber
JOURNAL      Unpublished (2000)
COMMENT      Contact: Wing RA
                Clemson University Genomics Institute
                Clemson University
                100 Jordan Hall, Clemson, SC 29634, USA
                Tel: 864 656 7288
                Fax: 864 656 4293
                Email: rwing@clemson.edu
                Seq primer: TAATACCACTCACTATAGGG
                High quality sequence start: 3
                High quality sequence stop: 791.
FEATURES
    Location/Qualifiers
        1..850
            /organism="Gossypium arboreum"
            /strain="AKA"
            /cultivar="8400"
            /db_xref="taxon:29729"
            /clone="GA_Ea0025E19f"
            /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
            /tissue_type="Fibers isolated from bolls harvested 7-10
            dpa"
            /lab_host="E. coli"

```


LOCUS BE237947 599 bp mRNA linear EST 11-JUL-2000
 DEFINITION 894039811.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
 Chlamydomonas reinhardtii cDNA, mRNA sequence.
 ACCESSION BE237947
 VERSION BE237947.1 GI:9027907
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaeae; Chlamydomonas.
 1 (bases 1 to 599)
 Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,
 McDermott, J. P., Sillflow, C., Stern, D. and Surzycki, R.
 Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants; project phase 2
 Unpublished (2000)
 CONTACT: Elizabeth H. Harris
 DCMB Box 91000
 Duke University
 Durham, NC 27708-1000, USA
 Tel: 919 613 8164
 Fax: 919 613 8177
 Email: chlamy@duke.edu

FEATURES
 source
 1..599
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
 II"
 /note="vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
 XhoI; This library, constructed by John Davies and Jeffrey
 McDermott, combines cDNAs from CC-1690 cells grown to
 mid-log phase in TAP (acetate-containing) medium in the
 light, TAP medium in the dark, HS (minimal) medium in
 ambient levels of CO₂ and HS medium bubbled with 5% CO₂.
 PolyA mRNA was purified from each sample, pooled and cDNA
 synthesized. The cDNA was directionally cloned into lambda
 Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
 pBluescript II SK- plasmids were excised from the lambda
 Zap clones by superinfection with ExAssist (Stratagene)
 phage. The library was normalized using method 4 described
 in Bonaldo et al (1996) Genome Research 6: 791-806."
 BASE COUNT 96 a 193 C 214 G 96 T
 ORIGIN

Alignment Scores:
 Pred. No.: 4.01e-24 Length: 599
 Score: 269.00 Matches: 61
 Percent Similarity: 60.14% Conservative: 28
 Best Local Similarity: 41.22% Mismatches: 53
 Query Match: 17.72% Indels: 6
 DB: 10 Gaps: 3
 US-09-308-397-2 (1-306) x BE237947 (1-599)

QY 83 Pro-----AspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuVal 98
 Db 395 GCCATTGACGCCATTGACGTGGCGCTGTCGTGGCGAGTACACGGCCTGGCC 454
 QY 99 AlaserGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyr 118
 Db 455 TTGCTGTGGTCATGAGCTTCAGAGCGGCTGCGGTGTAAGCTCGCGGCGAGTCC 514
 QY 119 MetGluGluAlaAlaProAlaaspSerGlyLysMetValAlaValLeuAsnThrProVal 138
 Db 515 ATGACGGCGGTGCTGACGCCAGCCAGCAGCATGTGTCCTGTCATTGGCTGGACAGC 574
 QY 139 GluValIleGluGluAlaCysGln 146
 Db 575 GCCAAGTGGCGGAGCTGTGCAAG 598

Search completed: June 11, 2003, 21:24:41
 Job time : 1437 secs